

**LAPORTAN PENELITIAN
PROGRAM STUDI**

**MODEL PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU
MENGUNAKAN METODE *NAÏVE BAYES***



Armansyah, M.Kom (Ketua)

Rakhmat Kurniawan. R, ST., M.Kom (Anggota)

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUMATERA UTARA
MEDAN**

2021

LEMBAR PENGESAHAN


1. a. Judul Penelitian : Model Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Naïve Bayes*
- b. Klaster Penelitian : Penelitian Program Studi
- c. Bidang Keilmuan : Ilmu Komputer
- d. Kategori : Kelompok
2. Peneliti 1 : Armansyah, M.Kom (Ketua)
3. ID Peneliti 1 : 200410840110000
4. Peneliti 2 : Rakhmat kurniawan. R, ST.,M.Kom (Anggota)
5. ID Peneliti 2 : 201603850113000
6. Unit Kerja : Fakultas Sain dan Teknologi, UINSU Medan
7. Waktu penelitian : 2 bulan di tahun 2021
8. Biaya Penelitian : Rp. 6.000.000,-
(Enam Juta Rupiah)

Medan, 6 Oktober 2021

Disahkan Oleh
Dekan
Fakultas Sain dan Teknologi


Dr. M. Syahnan, M.A
NIP.196609051991031000

Peneliti
Ketua


Armansyah, M.Kom
NIB. 1100000074

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertandatangan dibawah ini :

Nama : Armansyah
Jabatan : Dosen / Asisten Ahli / Gol. III.b
Unit Kerja : Fakultas Sain dan Teknologi, UIN SU Medan
Alamat : Jl. IAIN No. 1 Medan

dengan ini menyatakan bahwa :

1. Judul penelitian “Model Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Naïve Bayes*” merupakan karya orisinil saya.
2. Jika dikemudian hari ditemukan fakta bahwa judul, hasil, atau bagian dari laporan penelitian saya merupakan karya orang lain dan/atau plagiasi, maka saya akan bertanggungjawab mengembalikan 100% dana hibah penelitian yang telah saya terima, dan siap menerima sanksi sesuai ketentuan yang berlaku.

Demikian pernyataan ini dibuat untuk digunakan sebagaimana mestinya.

Medan, 6 Oktober 2021

Yang menyatakan,



Armansyah, M.Kom

NIB. 1100000074

ABSTRAK

Adanya ketidakselarasan jumlah lulusan mahasiswa dengan jumlah mahasiswa baru di setiap tahunnya menunjukkan pandangan yang kurang baik bagi Program Studi. Dari pantauan kami, ada kurang lebih 44 mahasiswa yang telah lulus di Oktober 2020 dari 949 mahasiswa mulai dari angkatan pertama 2015 sampai 2019. Ketidakselaran itu, perlu dicarikan solusi permasalahannya, yang kami menganggap solusi awal dari penyelesaian masalah itu adalah memodelkan prediksi kelulusan mahasiswa. Adapun model prediksi yang kami gunakan dalam penelitian ini adalah dengan metode klasifikasi *naïve bayes*. Data yang digunakan dalam penelitian kami ini sebanyak 44 data lulusan yang dibagi dalam 38 data untuk pelatihan dan 6 data lulusan untuk pengujian model prediksi menggunakan RapidMiner 7.6.0. Dari hasil pelatihan dan pengujian data yang terdiri dari 9 atribut dan 1 label/kelas, didapatkan kesimpulan bahwa akurasi model prediksi sebesar 100%.

Kata Kunci : Model Prediksi, Naïve Bayes, Kelulusan Tepat Waktu,
Klasifikasi, Pembelajaran Mesin

ABSTRACT

The discrepancy between the number of graduate students and the number of new students each year shows an unfavorable view for the Study Program. From our observations, there are approximately 44 students who have graduated in October 2020 from 949 students from the first batch of 2015 to 2019. This discrepancy requires a solution to the problem, which we consider the initial solution to solving the problem is to model predictions for student graduation. The prediction model that we use in this study is the naive Bayes classification method. The data used in this study are 44 graduate data which are divided into 38 data for training and 6 graduate data for testing prediction models using RapidMiner 7.6.0. From the results of training and testing of data consisting of 9 attributes and 1 label/class, it was concluded that the accuracy of the prediction model was 100%.

Keywords: Prediction Model, Nave Bayes, On Time Graduation,

Classification, Machine Learning

KATA PENGANTAR

Bismillahirrahmanirrahiim

Alhamdulillah, puji dan syukur kami sampaikan kepada Allah SWT. yang telah memberikan nikmat hidup kepada kami, yang dengan nikmat itu kami masih sempat menggunakan waktu, ilmu, serta pikiran dan tenaga untuk menyelesaikan laporan penelitian ini. Laporan ini merupakan laporan penelitian Program Studi, yang disusun untuk mendokumentasikan tahapan penelitian yang dilakukan selama penelitian terlaksana. Penelitian ini dilakukan oleh dosen Program Studi Ilmu Komputer dan dilaksanakan secara berkelompok, dengan tujuan untuk meningkatkan kemampuan dosen dalam melakukan penelitian.

Penelitian ini merupakan penelitian analisis berupa pemodelan prediksi kelulusan mahasiswa tepat waktu dan tidak tepat waktu (terlambat) dengan menggunakan metode klasifikasi *naïve bayes*. Dalam pelaksanaannya, analisis untuk model prediksi dalam penelitian ini menggunakan 9 atribut diantaranya jenis kelamin, Indeks Prestasi Semester 1 - Indeks Prestasi Semester 4, jumlah beban sks mahasiswa pada semester 4, masa tugas akhir dalam semester, masa studi mahasiswa dalam satuan tahun, dan jalur masuk perguruan tinggi, serta 1 atribut kelas kelulusan yang diwakili dengan label “tepat waktu” dan “tidak tepat waktu.”

Dari hasil pelatihan 38 data serta setelah dilakukan pengujian data lulusan yang berjumlah 6 data mahasiswa, menghasilkan akurasi pemodelan prediksi sebesar 100%. Dengan kata lain bahwa setelah pelatihan data, pengujian terhadap 6 data lulusan dengan perangkat lunak bantu RapidMiner 7.6.0 menunjukkan keluaran yang sama dan logis berupa keputusan bahwa ke-6 mahasiswa tersebut masuk dalam kategori kelulusan ‘tidak tepat waktu’. Secara detail penelitian ini disajikan dalam laporan penelitian yang dibagi

dalam 5 bab yaitu bab latar belakang, tinjauan pustaka, metodologi penelitian, analisis dan hasil serta penutup.

Dalam laporan ini, mungkin saja dijumpai kekurangan baik dalam penulisan maupun dalam isi pelaporan. Karenanya, kami mengharapkan pengembangan oleh pembaca atau pengguna laporan ini untuk kesempurnaan laporan ini. Kami mengharapkan laporan penelitian ini memberikan manfaat kepada pembaca, atau kepada pengguna laporan ini. Akhirnya, kami mengucapkan terimakasih atas kesempatan pembaca dan pengguna laporan ini, serta kami sampaikan selamat menulis dan selamat berkarya. Terima Kasih.

Medan, Oktober 2021

Tim Peneliti.



Armansyah, M.Kom
NIB. 1100000074

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN BEBAS PLAGIASI	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Rumusan Masalah	3
C. Tujuan Penelitian.....	4
D. Manfaat Penelitian.....	4
E. Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	6
A. <i>Machine Learning</i>	6
B. Pendekatan <i>Machine Learning</i>	7
C. Klasifikasi.....	9
D. Probabilitas.....	9
E. <i>Naïve Bayes</i>	10
F. <i>Confusion Matrix</i>	11
G. <i>Flowchart</i>	13
H. <i>RapidMiner</i>	14
I. Faktor Kelulusan	14
J. Penelitian Terkait	17

BAB III METODOLOGI PENELITIAN.....	24
A. Perencanaan	24
Kerangka Penelitian.....	25
Tempat dan Waktu Penelitian	26
Disain Penelitian.....	27
Objek Penelitian	28
Variabel Penelitian	28
Variabel Dependensi	28
Variabel Independen.....	28
B. Pelaksanaan	29
Teknik Pengumpulan Data	29
Pengolahan atau Analisis Data	29
Penyusunan Laporan	31
BAB IV ANALISIS DAN HASIL	32
A. Analisis Kebutuhan	32
B. Transformasi Data	39
C. Penyajian Data	41
D. Analisis dan Pemodelan <i>Naïve Bayes</i>	46
Menghitung Probabilitas Kelas	46
Menghitung Probabilitas Kejadian Atribut	47
Membandingkan Probabilitas Label/Kelas.....	50
E. Penerapan Model dengan <i>Naïve Bayes</i>	51
Model Prediksi dengan <i>Naïve Bayes</i>	51
Pengujian Data dengan Model.....	56
Akurasi Model	58
Temuan Penelitian	62
BAB V PENUTUP	63
A. Kesimpulan.....	63

B. Saran	63
DAFTAR REFERENSI.....	66

DAFTAR GAMBAR

Gambar 2.1 Antarmuka RapidMiner 7.6.0.....	14
Gambar 3.1 Kerangka Penelitian.....	26
Gambar 4.1 Antarmuka palatihan model prediksi dengan naïve bayes ...	30
Gambar 4.2 Nilai probabilitas atribut dari pelatihan dengan rapidminer..	52
Gambar 4.3 Grafik sebaran data lulusan berdasarkan jenis kelamin.....	53
Gambar 4.4 Distribusi statistik dari atribut kelas	54
Gambar 4.5 Distribusi atribut pelatihan	54
Gambar 4.6 Grafik kelas pelatihan data	55
Gambar 4.7 Antarmuka model untuk pengujian data.....	55
Gambar 4.8 Hasil prediksi dari pengujian data	56
Gambar 4.9 Akurasi Pelatihan Model	57
Gambar 4.10 Performance vector.....	58
Gambar 4.11 Hasil prediksi dari pengujian kedua	60
Gambar 4.12 Akurasi model NB dari pengujian kedu	60
Gambar 4.13 Performance vector.....	61
Gambar 4.14 Grafik klasifikasi pada ujicoba kedua.....	61

DAFTAR TABEL

Tabel 2.1 Confusion Matrix.....	12
Tabel 2.2 Ragam Notasi flowchart.....	13
Tabel 2.3 Penelitian Relevan.....	17
Tabel 3.1 Rencana Pelaksanaan.....	27
Tabel 4.1 Data Asli Kelulusan.....	32
Tabel 4.2 Data Pelatihan.....	36
Tabel 4.3 Data Pengujian.....	38
Tabel 4.4 Transformasi data pelatihan	39
Tabel 4.5 Transformasi data pengujian	41
Tabel 4.6 Kelulusan Berdasarkan Jenis Kelamin	42
Tabel 4.7 Kelulusan Berdasarkan IP Semester 1.....	42
Tabel 4.8 Kelulusan Berdasarkan IP Semester 2.....	43
Tabel 4.9 Kelulusan Berdasarkan IP Semester 3.....	43
Tabel 4.10 Kelulusan Berdasarkan IP Semester 4.....	43
Tabel 4.11 Kelulusan Berdasarkan Sks Semester 4	44
Tabel 4.12 Kelulusan Berdasarkan Lama Tugas Akhir.....	45
Tabel 4.13 Kelulusan Berdasarkan Masa Studi.....	45
Tabel 4.14 Kelulusan Berdasarkan Jalur Masuk Kuliah	46

BAB I

PENDAHULUAN

A. Latar Belakang

Data mahasiswa Program Studi Ilmu Komputer yang dihimpun dari Akademik Program Studi, berjumlah 949 mahasiswa dari angkatan 2015 sampai dengan angkatan 2019. Data itu belum termasuk mahasiswa baru angkatan 2020 dan 2021. Dari jumlah mahasiswa angkatan 2015 - 2019 tersebut, ada sebanyak 44 mahasiswa yang telah lulus studi. Jika melihat data kelulusan berdasarkan angkatannya, ada setidaknya 38 data lulusan angkatan 2015 dari total penerimaan diangkatan pertama 67 mahasiswa, selanjutnya 6 mahasiswa lulus dari 154 mahasiswa angkatan 2016 (tabel lulusan prodi., 2020). Sementara untuk mahasiswa angkatan 2017 sebanyak 253 mahasiswa masih dalam proses bimbingan tugas akhir sejak riset ini dilakukan.

Jumlah data kelulusan yang disampaikan pada paragraf awal tersebut, masih tergolong kecil jika dibandingkan dengan jumlah mahasiswa diangkatannya, terutama untuk mahasiswa angkatan 2016, tidak terkecuali pada mahasiswa angkatan 2017 yang sampai riset ini dilaksanakan, belum adanya data kelulusan pada angkatan yang dimaksud. Untuk mencapai kelulusan, persyaratan yang ditetapkan program studi Ilmu Komputer, adalah (1) telah menyelesaikan perkuliahan sebanyak 148 sks termasuk sks kerja praktik lapangan, kuliah kerja nyata, serta seminar proposal dan seminar hasil penelitian (skripsi), disamping itu mahasiswa yang diperbolehkan lulus (2) tidak memiliki nilai E, dan (3) nilai D tidak lebih dari 7 sks.

Persyaratan kelulusan yang ditetapkan seperti pada uraian paragraf dua, pada dasarnya merupakan persyaratan yang mudah untuk dicapai. Disamping itu, persyaratan kelulusan diatas masih bersifat umum, yang juga diterapkan pada hampir semua perguruan tinggi. Dengan kata lain,

pencapaian kelulusan mahasiswa tidak terlalu rumit untuk menempuhnya, sepanjang memang mahasiswa benar-benar mengikuti prosedur yang ditetapkan setiap perguruan tinggi, termasuk program studi Ilmu Komputer UIN Sumatera Utara Medan. Fakta dilapangan menunjukkan ketidakselarasan data kelulusan yang berbanding jauh dengan mahasiswa baru yang masuk disetiap tahunnya.

Ketidakselarasan itu dipandang suatu masalah yang perlu dicarikan solusinya. Untuk mencari solusi dari permasalahan tersebut, kami mencoba membuat model prediksi kelulusan tepat dan tidak tepat waktu. Hasil penelitian ini nantinya menjadi rujukan bagi pimpinan khususnya ketua program studi Ilmu Komputer UIN Sumatera Utara Medan dalam membuat kebijakan untuk mendorong tingkat lulusan yang wajar dimasa mendatang. Untuk memodelkan prediksinya, penelitian ini menggunakan metode klasifikasi *naïve bayes*. Metode klasifikasi *naïve bayes* ini sudah sering digunakan untuk aplikasi prediksi, diantaranya [1] yang menggunakan metode *naïve bayes* untuk memprediksi karyawan tetap, [2] yang menggunakan metode *naïve bayes* untuk klasifikasi kelulusan tepat waktu mahasiswa untuk prediksi kelulusan, dan masih banyak bukti-bukti riset yang menggunakan metode *naïve bayes* untuk prediksi.

Merujuk kepada uraian permasalahan yang telah diuraikan pada paragraf-paragraf diatas, serta mempertimbangkan perlunya melakukan prediksi kelulusan mahasiswa dalam membantu kebijakan pimpinan Program Studi dalam meningkatkan jumlah lulusan yang wajar dimasa mendatang, maka kami bermaksud melakukan riset dengan judul “*Model Prediksi Kelulusan Mahasiswa Tepat Waktu Dengan Metode Naïve Bayes*”.

B. Rumusan Masalah

Masalah inti dari penelitian ini adalah adanya kesenjangan nilai antara jumlah mahasiswa masuk dengan jumlah mahasiswa keluar lulus. Dengan kata lain, target kelulusan mahasiswa tepat waktu masih rendah atau perlu ditingkatkan. Untuk itu, rumusan masalah penelitian tersebut adalah dengan menjawab atau memenuhi pertanyaan rumusan permasalahan berikut yaitu :

1. Bagaimana meningkatkan jumlah lulusan sehingga didapatkan jumlah yang wajar antar jumlah angkatan masuk dengan jumlah mahasiswa keluar lulus?
2. Apa metode pemodelan prediksi kelulusan mahasiswa tepat waktu yang sesuai untuk merumuskan penyelesaian permasalahan ini?
3. Bagaimana memodelkan prediksi kelulusan mahasiswa tepat waktu dengan *naïve bayes* ?
4. Apakah metode *naïve bayes* dapat digunakan untuk memodelkan prediksi kelulusan mahasiswa tepat waktu, bagaimana melakukan pemodelannya?

Rumusan dari penelitian ini dapat dicapai dengan menjawab pertanyaan rumusan masalah diatas dengan melakukan langkah-langkah berikut :

1. Untuk meningkatkan jumlah kelulusan mahasiswa tepat waktu, adalah dengan membuat pemodelan prediksi kelulusan mahasiswa tepat waktu, dengan melihat faktor-faktor apa saja yang memengaruhi keterlambatan dan ketepatan lulusnya mahasiswa, sehingga pihak pengambil keputusan dapat menggunakan ini sebagai pengambilan kebijakan untuk meningkatkan jumlah kelulusan.

2. Untuk memodelkan prediksi kelulusan mahasiswa tepat waktu, digunakan metode *naïve bayes* dengan rujukan dari riset-riset yang telah ada.
3. Untuk memperoleh model prediksi kelulusan mahasiswa yang tepat, akan dilakukan analisis terhadap faktor-faktor penentu kelulusan tepat waktu dengan *naïve bayes*.
4. Untuk melihat bahwa pemilihan metode *naïve bayes* dapat digunakan dan cocok untuk penyelesaian permasalahan ini, akan dirujuk riset-riset terkait yang menunjukkan tingkat akurasi serta dengan menyesuaikan faktor-faktor penentunya.

C. Tujuan Penelitian

Mengacu kepada permasalahan penelitian yang telah disampaikan pada latar belakang, jelas bahwa penelitian ini dilakukan untuk tujuan sebagai berikut :

1. Memodelkan prediksi kelulusan mahasiswa tepat waktu dengan melibatkan faktor-faktor yang mendukung kelulusan mahasiswa tepat waktu,
2. Melakukan analisis faktor-faktor penentu kelulusan mahasiswa tepat waktu dengan metode *naïve bayes*,
3. Untuk mengetahui apakah metode *naïve bayes* dapat diterapkan untuk membuat pemodelan prediksi kelulusan melalui hasil analisis terhadap faktor-faktor yang memengaruhi kelulusan mahasiswa tepat waktu.

D. Manfaat Penelitian

Penelitian ini memberikan kontribusi yang banyak terutama kepada Program Studi Ilmu Komputer untuk membantu pihak pimpinan dalam membuat kebijakan dalam rangka meningkatkan jumlah kelulusan

mahasiswa yang tetap waktu, serta memberikan kontribusi kepada mahasiswa melalui kebijakan-kebijakan yang efektif untuk mendorong mahasiswa mencapai kelulusannya. Selain itu kontribusi lainnya adalah :

1. Untuk memperoleh pemodelan prediksi kelulusan mahasiswa tepat waktu,
2. Untuk mengetahui metode pemodelan prediksi kelulusan mahasiswa tepat waktu,
3. Untuk mengetahui prosedur analisis dengan metode *naïve bayes* dalam melakukan prediksi kelulusan mahasiswa,
4. Untuk mengetahui apakah metode *naïve bayes* dapat diterapkan pada permasalahan-permasalahan prediksi kelulusan mahasiswa tepat waktu melalui data yang memiliki faktor-faktor yang pemicu.

E. Batasan Masalah

Peneliti melihat bahwa permasalahan penelitian ini cukup luas jika dilihat dari faktor-faktor penentunya, dan pada aspek permasalahannya. Untuk itu, batasan masalah diberikan untuk memperkecil kegagalan riset ini. Yang diantaranya batasan masalah difokuskan pada :

1. Bahwa penelitian ini berupa model prediksi dengan melihat faktor-faktor berupa jenis kelamin, indeks prestasi yang diperoleh dari semester 1 hingga semester 4, beban SKS yang telah ditempuh mahasiswa pada semester iv, lama masa tugas akhir, masa studi mahasiswa, dan jalur masuk kuliah.
2. Pemodelan prediksi menggunakan metode *naïve bayes* sebagai metode analisisnya,
3. Data untuk analisisnya diperoleh dari Program Studi Ilmu Komputer UIN Sumatera Utara Medan, tempat dimana riset ini dilakukan.

BAB II

KAJIAN PUSTAKA

Untuk mencapai hasil penelitian, peneliti perlu mengkaji beberapa teori untuk mendukung terlaksananya penelitian ini sesuai dengan yang diharapkan. Penelitian ini mengacu kepada topik *machine learning* yang merupakan bagian dari bidang kecerdasan buatan, dan turunan dari bidang ilmu matematika. Secara rinci teori-teori yang digunakan dalam penelitian ini diuraikan pada subbab-subbab selanjutnya.

A. *Machine Learning*

Berbicara tentang *machine learning* (pembelajaran mesin), akan muncul berbagai pendapat tentangnya, tentang apa dan seperti apa *machine learning* itu. Jika melihat secara harfiah, yaitu diambil dari kata *machine* yang berarti mesin, dan *learning* yang berarti sedang belajar atau dapat juga dikatakan dengan pembelajaran. Bila arti dari dua kata itu disatukan, maka berarti mesin yang sedang belajar, atau pembelajaran mesin. Berarti bahwa, mesin melakukan pembelajaran. Mesin disini, bisa diwakili oleh komputer melalui program aplikasi yang melakukan proses pembelajaran terhadap data untuk tujuan tertentu.

Dalam prosesnya melakukan pembelajaran data, mesin (program komputer) melakukan tugas itu untuk tujuan seperti mengelompokkan (*clustering*), mengklasifikasi (*classification*), pengenalan pola (*patern recognition*), memprediksi (*prediction*), meramal (*forecasting*), dan tugas-tugas manusia yang lainnya yang mampu dilakukan oleh mesin sebagai bentuk kecerdasan buatan yang disematkan oleh manusia ke dalam mesin (komputer).

Mesin (komputer) itu melakukan tugas-tugas dalam melakukan pengelompokan, klasifikasi, prediksi dan peramalan, melalui data-data atau

dataset. Dataset tersebut, mungkin saja belum memiliki pola. Disinilah manusia membuat pola-pola data tertentu yang memungkinkan dapat dipahami oleh mesin komputer. Pada tahapan awal, manusia akan mengenalkan pola-pola data tersebut dengan algoritma tertentu yang sesuai dengan representasi data. Tahapan-tahapan yang disusun dalam rancangan algoritma itu akan dipandang sebagai pengalaman bagi mesin komputer, yang menjadi dasar baginya untuk mengenali pola-pola data lainnya di masa mendatang. Dengan pengalaman mengenali pola-pola data inilah indikasi mesin komputer dapat belajar. Yakni belajar melalui data, dan mempelajari data tersebut untuk membuat suatu kesimpulan dari tugasnya sebagai pembelajaran mesin. Uraian penjelasan tersebut, senada dengan definisi *machine learning* seperti yang disampaikan oleh Abhishek Vijayvargia (2018) dalam bukunya[3].

B. Pendekatan *Machine Learning*

Machine learning, tidak dengan begitu saja dapat mempelajari data, sama halnya dengan manusia, manusia memiliki beberapa teknik dan strategi dalam melakukan pembelajaran. Begitu juga dengan program komputer. *Machine learning*, menggunakan teknik pembelajaran Supervised learning, Unsupervised learning, dan, Reinforcement learning[4].

1. *Supervised Learning*

Supervised learning merupakan pendekatan pembelajaran *machine learning*, dimana pendekatan ini disebut juga dengan pembelajaran terbimbing atau terawasi. Teknik pembelajaran ini ibarat pembelajaran yang dilakukan oleh seorang murid dengan gurunya. Dimana murid melakukan tugas yang diarahkan kepada guru, dan guru sebagai validasi terhadap tugas muridnya. Guru akan memberikan validasi benar atau kurang tepat untuk tugas yang diberikan. Untuk tugas yang dipandang sebagai tugas yang benar,

maka ini mengindikasikan bahwa murid telah mampu dalam mengerjakan tugasnya. Sementara untuk tugas yang kurang tepat, akan diarahkan untuk mengulang tugas muridnya tanpa melakukannya dari semua. Artinya, untuk mencocokkan antara tugas murid dengan harapan guru, guru sebagai pembimbing hanya memberikan catatan pada bagian yang belum sesuai. Dalam pembelajaran mesin, pendekatan pembelajaran terbimbing akan mencoba menemukan data yang tepat, untuk menemukan data yang tepat ini, bobot dibutuhkan untuk dapat mengenali pola yang diharapkan. Disini, mesin membutuhkan seperangkat data input untuk pengolahan data, dan data target sebagai pembimbing.

2. *Unsupervised Learning*

Unsupervised Learning merupakan kebalikan dari Supervised Learning. Yakni melakukan pendekatan pembelajaran yang tidak diawasi atau tidak dibimbing. Unsupervised Learning paling cocok ketika masalah membutuhkan sejumlah besar data yang tidak berlabel. Misalnya, aplikasi media sosial, seperti Twitter, Instagram, Snapchat, dan sebagainya semuanya memiliki data yang tidak berlabel dalam jumlah besar. Dengan besarnya data yang dimaksud diatas, menjadikannya tidak dapat diklasifikasikan langsung. Umumnya, perlu tahapan pengelompokan (clustering) terlebih dahulu sebelum mengklasifikasikan data ke kategori tertentu.

3. *Reinforcement Learning*

Reinforcement learning merupakan model pembelajaran tingkah laku. Yakni melakukan pendekatan belajar observasi. Sistem belajar pada jenis ini melalui mekanisme *try and error*. Untuk hasil yang dipandang baik atau benar, akan disimpan dan menjadi penguat pada pelatihan data selanjutnya.

C. Klasifikasi

Klasifikasi adalah proses menemukan model (atau fungsi) yang menggambarkan dan membedakan kelas atau konsep data. Model diturunkan berdasarkan analisis sekumpulan data pelatihan (yaitu, objek data yang label kelasnya diketahui). Model yang digunakan untuk memprediksi label kelas objek yang label kelasnya tidak diketahui[5].

Tahapan klasifikasi data dibagi dalam 2 langkah proses. Langkah pertama adalah tahapan *learning* yang disebut sebagai *fase training*. Fase ini merupakan tahapan algoritma klasifikasi untuk membentuk aturan klasifikasi data yang diambil melalui analisi data *training*. Tahap kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari aturan klasifikasi.

D. Probabilitas

Fitria Hidayanti (2017) dalam buku yang berjudul *Probabilitas dan Statistik: Aplikasi Probabilitasi* mengemukakan probabilitas merupakan ilmu ketidakpastian [6]. Dimana ketidakpastian menekankan pada suatu kejadian mungkin saja terjadi tidak dari seluruh kejadian (seluruh kelas) yang ada. Nilai suatu probabilitas diwakili oleh nilai antara 0 dan 1 atau dalam persentase. Nilai probabilitas 1 mewakili suatu kejadian memungkinkan pasti terjadi atau mungkin sudah terjadi. Sedangkan nilai probabilitas 0 mewakili kejadian yang mustahil atau tidak mungkin terjadi [7]. Nilai probabilitas ditentukan dengan persamaan :

$$P = \frac{X}{N} \quad (2.1)$$

Dimana,

P, merupakan probabilitas kejadian

X, merupakan jumlah kejadian suatu kelas

N, merupakan total kejadian dari suatu kelas

E. Naïve Bayes

Naive Bayes adalah teknik probabilistik untuk membangun pengklasifikasi. Asumsi karakteristik dari pengklasifikasi *naïve bayes* adalah untuk mempertimbangkan bahwa nilai fitur tertentu tidak tergantung pada nilai fitur lainnya, mengingat variabel kelas. Misalnya, sebuah ponsel dapat dianggap cerdas jika memiliki layar sentuh, fasilitas internet, kamera yang bagus, dan lain-lain. Meskipun semua fitur ini saling bergantung satu sama lain, fitur-fitur tersebut berkontribusi secara independen terhadap kemungkinan bahwa ponsel tersebut adalah ponsel pintar.

Dalam klasifikasi Bayesian, hal yang paling penting menemukan probabilitas *posterior* yaitu probabilitas label yang diberikan beberapa fitur yang diamati, (L). Dengan bantuan teorema Bayes, kita dapat menyatakannya dalam bentuk formula berikut ini :

$$P(H|X) = \frac{P(H).P(X|H)}{P(X)} \quad (2.2)$$

Dimana,

X, merupakan data dengan kelas yang belum diketahui

H, merupakan hipotesis data yang merupakan suatu kelas spesifik

$P(H|X)$, merupakan probabilitas hipotesis H berdasarkan pada kondisi X (posterior probabilitas)

$P(H)$, merupakan probabilitas hipotesis H (prior probabilitas)

$P(X|H)$, merupakan probabilitas X berdasarkan pada kondisi pada hipotesis H, dan

$P(X)$, merupakan probabilitas X

Atau dapat boleh juga dipandang :

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (2.3)$$

Nilai *evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Sementara klasifikasi data yang bersifat kontinyu dapat menggunakan formula densitas gaussian dengan formula dibawah ini :

$$P(X_i = x_i | Y = Y_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2(\sigma)^2}} \quad (2.3)$$

Dimana,

P : adalah peluang

X_i : Atribut ke-i

x_i : Nilai atribut ke-i

Y : Kelas yang dicari

μ : Merupakan rerata dari seluruh atribut

σ : Standar deviasi

F. Confusion Matrix

Confusion Matrix merupakan prosedur metode yang digunakan untuk memecahkan kebingungan antara kebenaran metode prediksi dengan keadaan yang real. Dengan kata lain, disini metode pridiksi perlu melalui uji validitas sehinga keluaran dari metode prediksi bersifat benar atau setidaknya dapat dipercaya.

Evaluasi dengan metode ini umumnya mengukur tingkat akurasi, presisi, dan sensitifitas dari kinerja prediksi [8]. Confusion matrix tidak lain adalah sebuah tabel dengan dua dimensi yakni “Aktual” dan “Diprediksi” dan selanjutnya, kedua dimensi menyisir pada “True Positive (TP)”, “True Negative (TN)”, “False Positive (FP)”, “False Negative (FN)” seperti yang ditunjukkan tabel di bawah ini :

Tabel 2.1 *Confusion Matrix*

		Nilai Kebenaran	
		1	0
Nilai Prediksi	1	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	0	<i>True Negative (TN)</i>	<i>False Negative (FN)</i>

Dari tabel diatas akan dilihat nilai kebenaran (nilai yang sebenarnya) dan nilai prediksi dari kinerja metode yang digunakan. Dimana kebenaran dan prediksi diwakili dengan bilangan 1 dan 0. Bilangan 1 mewakili kebenaran, dan bilangan 0 mewakili kesalahan antara keadaan sebenarnya dengan prediksinya. Secara rinci penjelasan tabel 2.1 dijelaskan dibawah ini :

- *True Positives (TP)*: Keadaan dimana kelas sebenarnya dan kelas prediksi dari titik data adalah 1.
- *True Negatives (TN)*: Keadaan dimana ketika kelas sebenarnya dan kelas prediksi dari titik data adalah 0.
- *False Positive (FP)*: Keadaan dimana kelas sebenarnya dari titik data adalah 0 dan kelas titik data yang diprediksi adalah 1.
- *False Negative (FN)*: Keadaan dimana kelas sebenarnya dari titik data adalah 1 dan kelas prediksi titik data adalah 0.


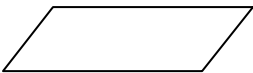
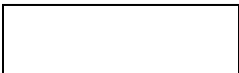
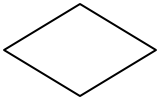
Dari tabel *confusion matrix* diatas dapat dihitung tingkat akurasi dari kinerja metodenya dengan menggunakan formula sebagai berikut :



$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} * 100\% \quad (2.7)$$

G. *Flowchart*

Diagram alur atau yang disebut dengan *flowchart* merupakan notasi yang menggambarkan alur atau tahapan suatu proses. Dimana, suatu alur proses disimbolkan (dinotasikan) dengan bentuk-bentuk simbol gambar. Biasanya yang dimaksud dengan notasi ini berupa tahapan-tahapan seperti memulai proses, memproses sesuatu yang bersifat berulang, proses pengambilan keputusan, masukan, keluaran dan lain sebagainya. Adapun notasi simbol diagram alur ini diantaranya ditunjukkan pada tabel berikut :

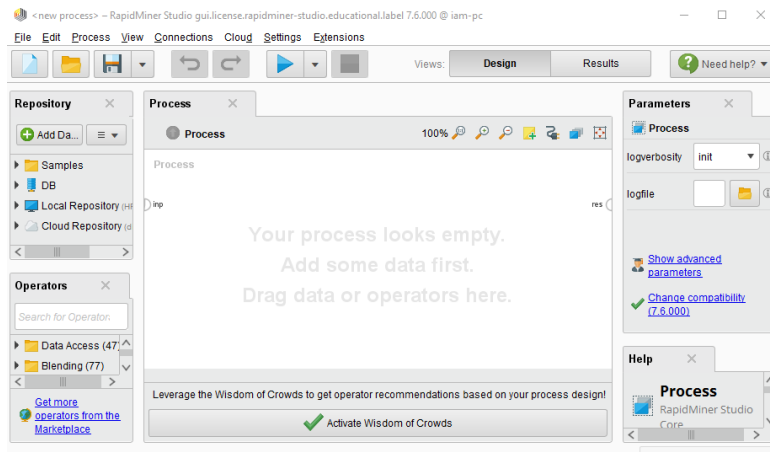
Tabel 2.2 Ragam Notasi *flowchart*

Notasi	Kegunaan
	Merupakan notasi diagram alur yang digunakan untuk menggambarkan awal dan akhir proses
	Merupakan notasi diagram alur yang digunakan untuk menggambarkan suatu masukan yang tidak terdefinisikan alat masukannya.
	Merupakan notasi diagram alur yang menggambarkan proses, biasanya proses pembacaan data, komputasi oleh sistem, dan proses-proses lainnya.
	Merupakan notasi diagram alur yang menggambarkan pengambilan keputusan, dimana keputusan yang dimaksud disini bisa berupa keputusan untuk mengulang sesuatu atau memilih sesuatu.

Notasi	Kegunaan
	Merupakan notasi diagram alur yang berfungsi menghubungkan diagram alur lain yang berada di halaman yang sama.
	Merupakan notasi diagram alur yang berfungsi menghubungkan diagram alur di halaman lain.

H. RapidMiner

RapidMiner merupakan perangkat lunak bantu untuk melakukan analisis terhadap *data mining*, *text mining*, untuk melakukan klasifikasi data dan prediksi [9]. RapidMiner menyediakan berbagai fasilitas untuk analisis data, apakah untuk tujuan klusterisasi, klasifikasi, prediksi atau untuk tujuan lainnya yang menerapkan pendekatan pembelajaran mesin (*machine learning*).



Gambar 2.1 Antarmuka RapidMiner 7.6.0

I. Faktor Kelulusan

Jodi Letkiewicz, dkk. (2015) dalam laporan riset mereka menyampaikan bahwa mahasiswa yang banyak terlibat dengan lingkungan kampus, dalam

hal ini adalah tinggal dilingkungan kampus merupakan faktor penentu kelulusan mahasiswa tepat waktu. Mereka beranggapan bahwa bilamana mahasiswa berada dekat dengan kampusnya, memungkinkan mereka akan lebih fokus belajar, serta mengakses berbagai fasilitas dan prasarana belajar. Mereka merujuk kepada pernyataan Rektor Universitas Gordon Gee dari laporan riset Bush dan Pyle tahun 2011 dengan menyimpulkan bahwa memperluas area tempat tinggal mahasiswa di lingkungan kampus akan mengarah kepada “tingkat retensi dan kelulusan yang lebih tinggi, nilai yang lebih baik, peningkatan keterlibatan dalam kegiatan kampus, dan lebih banyak pendampingan di antara siswa” [10].

Pendapat Jodi bisa saja benar, namun itu hanya sedikit dari banyak faktor. Dan pada umumnya faktor-faktor penentu bisa bergantung kepada pola pikir dan keadaan masyarakat dari setiap negara. Sementara penulis sepakat dengan pendapat R.A. Pertiwi [11] yang mengatakan bahwa faktor kelulusan dapat ditentukan dengan faktor belajar. Maksudnya disini adalah, dengan semangat belajar yang tinggi serta mengeksplorasi pengetahuan yang sesuai dengan bidang yang sedang diambil, apakah belajar mandiri maupun belajar bersama dengan sesama teman, baik diskusi maupun individu, serta belajar melalui sumber-sumber referensi maupun belajar melalui bimbingan di lembaga kursus dan sejenisnya, akan meningkatkan kompetensi berpikir dan kompetensi memecahkan masalah. Dengan meningkatnya kemampuan berfikir dan memecahkan masalah, akan mendorong mahasiswa cepat lulus karena kemampuannya dalam menghadapi berbagai penugasan dalam kegiatan perkuliahan.

Dalam kbbi.web.id kata lulus diartikan dengan, dapat masuk, dapat lepas, terperosok, dan berhasil. Dalam konteks pendidikan, “berhasil” merupakan kata yang tepat untuk mendefinisikan seorang mahasiswa yang keluar dari perguruan tinggi karena telah melalui proses pendidikan. Dengan

kata lain seorang mahasiswa dibiarkan bebas untuk keluar dari perguruan tinggi dengan gelar sarjana karena telah menyelesaikan semua tugas-tugas perkuliahan, baik ketika belajar pada mata kuliah umum, dasar kejuruan, mata kuliah konsentrasi hingga tugas akhir.

Kelulusan setiap mahasiswa, untuk setiap perguruan tinggi memiliki aturan tersendiri, secara umum adalah telah menyelesaikan beban mata kuliah. Kelulusan di Program Studi Ilmu komputer UIN Sumatera Utara Medan, menetapkan beban mata kuliah sebesar 144 Sistem Kredit Semester (SKS), termasuk diantaranya Kerja Praktik Lapangan, Kuliah Kerja Nyata (KKN), seminar proposal dan seminar hasil (skripsi) [12]. Disamping itu, Program Studi Ilmu Komputer memberlakukan prinsip kelulusan yang ditekankan pada Peraturan Pemerintah. Program Studi Ilmu Komputer, merujuk kepada Peraturan Pemerintah No. 19 Tahun 2005 yang mengatur Standar Nasional Pendidikan, dimana kelulusan mahasiswa harus memenuhi kualifikasi kemampuan lulusan dengan kriteria memiliki kemampuan sikap, kemampuan pengetahuan, dan kemampuan keterampilan.

Ketiga kemampuan (sikap, pengetahuan, dan keterampilan) tersebut akan terbentuk pada individu mahasiswa melalui proses pembimbingan oleh pembimbing mata kuliah, kebijakan-kebijakan yang berlaku pada perguruan tinggi, serta lingkungan dan kegiatan ekstrakurikuler, serta sarana dan prasarana yang tersedia bilamana sarana dan prasarana tersebut digunakan dengan semaksimal mungkin. Pencapaian kemampuan tersebut tidak sebatas memberikan kemudahan dalam mengerjakan tugas-tugas kuliah, tetapi juga memberikan dorongan kepada mahasiswa untuk mencapai kelulusan tepat waktu, atau bahkan kelulusan lebih awal.

Secara normal, kelulusan sarjana Program Studi Ilmu Komputer adalah 8 semester atau 4 tahun pendidikan. Namun, tidak menutup kemungkinan dapat ditempuh sampai 3,5 tahun masa studi. Kelulusan

mahasiswa yang dicapai selama 4 tahun, dipandang sebagai mahasiswa lulus tepat waktu. Untuk mahasiswa yang lulus lebih dari 4 tahun dipandang sebagai kelulusan yang terlambat. Sementara untuk mahasiswa yang lulus dibawah 4 tahun, merupakan kelulusan yang cepat. Semua mahasiswa memiliki peluang untuk lulus tepat waktu, terlambat, atau lebih cepat bergantung kepada kemampuan sikap, pengetahuan dan keterampilan yang dimilikinya.

J. Penelitian Terkait

Penelitian ini merupakan pemodelan prediksi dengan menggunakan metode klasifikasi, yang merupakan penerapan pembelajaran mesin (*machine learning*). Penelitian yang dengan terkait *machine learning* itu sendiri sudah sering dilakukan oleh kalangan peneliti. Secara rinci penelitian yang relevan dengan penelitian kami disini diantaranya sebagai berikut pada tabel :

Tabel 2.3 Penelitian Relevan

NO.	PENELITI	JUDUL	KESIMPULAN
1	Azis Zed Ali Muhammad Bajabir (2018)	Penerapan metode <i>naive bayes</i> untuk prediksi menentukan karyawan tetap Pada PT. YSP <i>industries</i> Indonesia	Penelitian ini bertujuan untuk menerapkan <i>naive bayes</i> dalam melakukan prediksi karyawan yang layak menjadi karyawan tetap. Dari hasil penelitian ini berupa kesimpulan bahwa metode klasifikasi <i>naive bayes</i> dapat melakukan prediksi berdasarkan atribut absensi, kepribadian, dan prestasi, dengan tingkat akurasi 94% menggunakan perangkat bantu RapidMiner.
2	Devi Heryana (2019)	Data mining untuk memprediksi	Hasil penelitian pengujian menggunakan RapidMiner

NO.	PENELITI	JUDUL	KESIMPULAN
		kelulusan mahasiswa Pendidikan matematika uin raden intan lampung Menggunakan <i>naive bayes</i> .	5.3 dengan data <i>training</i> sebanyak 51 data diperoleh keakurasian sebesar 74,67%. Sementara data <i>testing</i> sebanyak 184 mahasiswa sebagai data uji didapatkan hasil bahwa mahasiswa yang akan lulus tepat waktu sebanyak 42 mahasiswa atau sekitar 22,8% dari jumlah data <i>testing</i> dan mahasiswa yang tidak tepat waktu sebanyak 142 mahasiswa atau sekitar 77,2%.
3	Erene Fajrila (2018)	Perbandingan klasifikasi ketepatan waktu Kelulusan mahasiswa menggunakan regresi Logistik biner dan <i>naive bayes classifier</i>	Penelitian ini bertujuan untuk memprediksi kelulusan mahasiswa tepat atau tidak waktu dengan menerapkan pengklasifikasian. Klasifikasi menggunakan Regresi Logistik Biner menggunakan variabel yang signifikan tersebut mendapatkan akurasi sebesar 82,29%, sedangkan hasil klasifikasi menggunakan Algoritma Naïve Bayes sebesar 89,10%. Hal ini dapat dikatakan pengklasifikasian menggunakan Naïve Bayes Classifier lebih baik dibandingkan Regresi Logistik Biner.
4	Susi Mashlahah	Prediksi Kelulusan Mahasiswa	Prediksi kelulusan mahasiswa tahun 2005-

NO.	PENELITI	JUDUL	KESIMPULAN
	(2013)	Menggunakan Metode Decision Tree	2008, menghasilkan akurasi sebesar 82,79%, dengan atribut data pelatihnnya daerah asal, jenis sekolah, jalur masuk, pengalaman pesantren, indeks prestasi kumulatif, indeks prestasi semester 1-5.
5	Fahmi Izhari (2019)	Analisis algoritma jaringan saraf tiruan Backpropagation terhadap kemampuan siswa Berdasarkan aspek kognitif Tesis	Prediksi yang dilakukan dalam penelitian ini menggunakan model Jaringan Syaraf Tiruan (JST) Backpropagation, untuk memprediksi kemampuan siswa berdasarkan aspek kognitif. Pelatihan dan pengujian JST terhadap data memperoleh akurasi sebesar 82.93% dan 85.37% untuk dua node <i>output</i> .
6	Lila Dini Utami (2015)	Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma <i>Naive Bayes</i>	Penerapan <i>naive bayes</i> dalam melakukan prediksi, terhadap pengolahan data yang dilakukan dalam 3 tahap, yakni <i>naive bayes</i> , <i>naive bayes</i> dan <i>information gain</i> , dan <i>naive bayes</i> , <i>information gain</i> , dan <i>adaboost</i> . Hasilnya, prediksi dengan <i>naive bayes</i> saja memperoleh akurasi hanya mencapai 70% dan $AUC=0,500$. Begitu juga dengan prediksi menggunakan kombinasi <i>naive bayes</i> dan <i>information gain</i> , akurasinya juga 70% dengan $AUC=0,500$, namun

NO.	PENELITI	JUDUL	KESIMPULAN
			prediksi dengan kombinasi antara naïve bayes, information gain disertai dengan metode adaboost, akurasi meningkat 29,5% menjadi 99,5% dan AUC=0,995.
7	Naziah Amalia, Shaufiah, Siti Sa'adah	Penerapan Teknik Data Mining Untuk Klasifikasi Ketepatan Waktu Lulus Mahasiswa Teknik Informatika Universitas Telkom Menggunakan Algoritma <i>Naive Bayes Classifier</i>	Penelitian ini penelitian prediksi dengan menggunakan atribut kota asal, jumlah SKS, IPK, berapa kali mengambil mata kuliah TA menggunakan <i>naive bayes</i> . Hasil pengujian diperoleh akurasi sebesar 91%.
8	Rindy Anggun Pertiwi, Indahwati, Farit Mohamad Afendi (2013)	Analisis chaid untuk identifikasi ketepatan waktu lulus Berdasarkan karakteristik mahasiswa	Penelitian ini berupa klasifikasi kelulusan mahasiswa dengan menggunakan diagram pohon CHAID. Dari analisis yang dilakukan, penelitian ini menyimpulkan bahwa faktor utama mahasiswa lulus adalah yang memiliki IPK $\geq 3,39$ dari fakultas ekonomi manajemen, sementara IPK 2,77 dari fakultas MIPA. Berdasarkan dari itu, prediksi kelulusan mahasiswa dapat dilakukan berdasarkan klasifikasi IPK.
9	Ade Ricky Rozzaqi (2015)	<i>Naïve Bayes</i> dan Filtering Feature Selection Information Gain untuk Prediksi Ketepatan Kelulusan Mahasiswa	Penelitian ini bertujuan untuk memprediksi kelulusan mahasiswa berdasarkan program studi menggunakan kombinasi metode <i>naive bayes</i> dan <i>information gain</i> , dengan

NO.	PENELITI	JUDUL	KESIMPULAN
			atribut yang digunakan adalah beban sks semester 1-6, dan ipk semester 1-6. Dari hasil penelitian menunjukkan bahwa nilai akurasi tertinggi diperoleh algoritma <i>Naive Bayes</i> dan algoritma <i>Fitur Selection Information Gain</i> dengan akurasi 89,79 %, sementara dengan metode <i>naive bayes</i> saja memperoleh 83% akurasi prediksinya.
10	Syarli (2016)	Metode <i>Naive Bayes</i> Untuk Prediksi Kelulusa (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)	Penelitian ini bertujuan untuk menerapkan metode <i>naive bayes</i> untuk memprediksi kelulusan pilihan mahasiswa ke perguruan tinggi. Dengan atribut pelatihan seperti Prodi., Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata. Hasil prediksi yang dilakukan dengan WEKA menunjukkan akurasi sebesar 94%.
11	Adrian Satria Putra (2018)	Klasifikasi Satus Gizi Balita menggunakan <i>Naive Bayes</i>	Penelitian ini bertujuan untuk mengklasifikasikan status gizi bayi berdasarkan atribut usia, berat badan, tinggi badan, gaji orangtua dengan menggunakan <i>naive bayes</i> . Hasil penelitian ini menunjukkan 93,1% akurasi pelatihan dan pengujian, yang dapat dijadikan sebagai aplikasi prediksi.
12	Roy Andri, MS. Hendriyawan	Implementasi metode <i>Naive Bayes Classifier</i> untuk	Penelitian ini bertujuan untuk menentukan prakiraan cuaca menggunakan metode

NO.	PENELITI	JUDUL	KESIMPULAN
	Ahmad (2019)	Prakiraan cuaca	<i>naive bayes</i> berdasarkan kecepatan angin, suhu, tekanan, curah hujan, dan fenomena atmosfer. Hasil pelatihan data dan pengujiannya menggunakan <i>naive bayes</i> diperoleh akurasi sebesar 82.67% dan galat sebesar 17.33% dengan menggunakan pengujian confusion matrix. Sementara pada pengujian Beta dengan 4 skala likeart menggunakan 30 responden diperoleh 38% menyatakan sangat setuju, 56.6% menyatakan setuju, 4.67% menyatakan tidak setuju dan 0.67% menyatakan sangat tidak setuju.
13	Nichyta Dian Ramadhani (2019)	Analisis Kinerja Algoritma <i>Naive Bayes Classifier</i> dengan Seleksi Fitur <i>Chi-Square</i> dan <i>Confix Stripping Stemmer</i> Pada Sistem Klasifikasi Terjemahan Hadits	Penelitian ini bertujuan untuk mengklasifikasi hadits berdasarkan terjemahannya kedalam kelas sebagai hadits berisi anjuran, larangan, dan informasi. Penelitian ini dapat digunakan untuk memprediksi keshohihan hadits yang diriwayatkan oleh Imam Bukhori. Penelitian ini menerapkan kombinasi algoritma <i>Naive Bayes Classifier</i> dengan <i>Chi-Square</i> dan <i>Confix Stripping Stemmer</i> , yang dalam pelatihan serta pengujian datanya menghasilkan klasifikasi

NO.	PENELITI	JUDUL	KESIMPULAN
			dengan akurasi berturut-turut 84,44%, 80,19%, 76,67% dan 76,21% dari 4 kali percobaan.

BAB III

METODOLOGI PENELITIAN

Metode diartikan sebagai cara yang tepat untuk melakukan sesuatu. Arti kata metode itu, sering disandingkan dengan kata “*logos*” yang berarti ilmu atau pengetahuan. Kedua kata tersebut menjadi kombinasi kata “metodologi” yang dimaknai dengan, cara melakukan sesuatu dengan menggunakan pikiran secara seksama untuk mencapai suatu tujuan. Menurut para ahli, satu diantaranya menurut Hidayat dan Sedarmayanti (2002, 25) di dalam artikel yang ditulis oleh Syafnidawaty dalam raharja.ac.id, merupakan pengkajian terhadap langkah-langkah dalam menggunakan sebuah metode. Sementara, masih merujuk kepada artikel Syafnidawaty, mengemukakan secara teknis tentang metode-metode yang digunakan dalam penelitian.

Penelitian ini merupakan penelitian berupa pemodelan untuk memprediksi kelulusan mahasiswa tepat waktu dengan menggunakan metode klasifikasi *naive bayes*. Adapun langkah-langkah penelitian ini disajikan dalam beberapa bagian yakni, perencanaan, pelaksanaan, dan penyusunan laporan.

A. Perencanaan

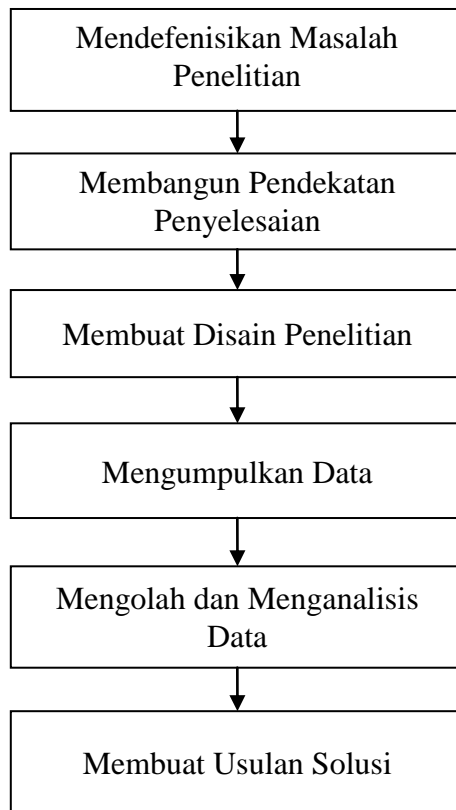
Tahapan ini merupakan tahapan awal dalam penelitian yang kami lakukan. Dimana tahapan ini mengacu kepada permasalahan umum yang berasal dari fenomena yang terjadi khususnya pada Program Studi Ilmu Komputer. Adapun fenomena yang diamati selama ini adalah, adanya kesenjangan nilai antara mahasiswa baru dengan angka kelulusan tepat waktu mahasiswa. Hal ini, dipandang sebagai masalah yang perlu diselesaikan dengan segera.

Dengan melihat fenomena yang terjadi, tahapan perencanaan pada penelitian ini dilakukan dengan beberapa kegiatan seperti :

- 1) Mendata kelulusan mahasiswa dan memberikan kesimpulan awal berupa penyesuaian antara data mahasiswa aktif dengan mahasiswa yang telah lulus,
- 2) Mengkaji fenomena dengan merujuk kepada penelitian terdahulu untuk menemukan metode apa yang digunakan dalam penelitian,
- 3) Merumuskan penyelesaian permasalahan dengan menggunakan metode yang dipandang sesuai, berdasarkan kepada riset-riset yang pernah dilakukan,
- 4) Menyusun kerangka penelitian,
- 5) Menentukan tempat dan waktu penelitian

Kerangka Penelitian

Menurut ahli, seperti Sugiono (2017) dalam artikel penelitian pada dosensosiologi.com, mengatakan bahwa kerangka penelitian merupakan alur berpikir dengan menerapkan berbagai model konseptual tentang bagaimana teori berhubungan dengan faktor yang telah diidentifikasi sebagai masalah dalam topik penelitian dengan susunan yang sistematis. Dengan melihat pengertian itu, maka susunan kerangka penelitian ini dapat dilihat pada bagan berikut :



Gambar 3.1 Kerangka Penelitian

Tempat dan Waktu Penelitian

Adanya kesenjangan antara jumlah mahasiswa aktif dengan mahasiswa keluar lulus, dipandang sebagai suatu masalah. Fenomena ini terjadi pada Program Studi Ilmu Komputer UIN Sumatera Utara Medan. Dengan permasalahan itu, penelitian berupa penentuan model prediksi dilakukan di unit yang telah disebutkan diawal, yang bertempat di alamat, Jl. IAIN No. 1 Medan, dan dilaksanakan pada Agustus - Oktober semester gasal tahun akademik 2021/2022. Secara rinci pelaksanaan penelitian ini dijelaskan pada tabel berikut :

Tabel 3.1 Rencana Pelaksanaan

NO	KEGIATAN	Agustus		September				Oktober	
		Mg 3	Mg 4	Mg 1	Mg 2	Mg 3	Mg 4	Mg 1	Mg 2
1	Pengajuan proposal								
2	Seminar proposal								
3	Pelaksanaan penelitian :								
	a. Mengumpulkan data latih								
	b. Mengolahan data dan analisis								
4	Penyusunan laporan								
5	Penyampaian laporan dan publikasi								

Disain Penelitian

Penelitian ini berupa pemodelan prediksi kelulusan mahasiswa tepat waktu. Dimana metode penelitiannya adalah kuantitatif, yang digunakan sebagai pendekatan penelitian yang berlandaskan pada filsafat positivisme digunakan untuk meneliti pada populasi atau sampel tertentu, pengumpulan data menggunakan instrument penelitian, analisis data bersifat kuantitatif dengan tujuan untuk menguji hipotesis yang telah ditetapkan (Sugiono, 2016:11) dalam artikel[13]. Masih dari sumber artikel yang sama, menurut

Moh. Sidik Priadana, penelitian kuantitatif menekankan pada variabel penelitian dengan angka dan melakukan analisis data dengan prosedur statistik.

Pendekatan kuantitatif pada penelitian ini mengacu kepada metode klasifikasi *naive bayes* yang digunakan untuk menganalisis data untuk menghasilkan keluaran berupa klasifikasi. Sementara analisis data dilakukan terhadap objek penelitian dan variabel penelitian.

Objek Penelitian

Objek penelitian yang digunakan disini adalah kinerja akademis mahasiswa Program Studi Ilmu Komputer UIN Sumatera Utara Medan yang menjadi tolok ukur terhadap kelulusan mahasiswa.

Variabel Penelitian

Sesuai dengan pendapat Moh. Sidik Priadana diatas, yang menyatakan bahwa variabel merupakan poin penting dalam penelitian kuantitatif. Dimana variabel penelitian terdiri dari variabel dependen dan variabel independen. Variabel dependen merupakan variabel yang memungkinkan terpengaruh dengan variabel lain. Sementara independen, bersifat sebaliknya yaitu tidak bergantung kepada variabel lain.

Variabel Dependen

Variabel dependen yang dimaksudkan pada penelitian ini adalah variabel kelulusan mahasiswa, yakni kelulusan mahasiswa tepat waktu, dan terlambat (tidak tepat waktu). Dimana indikator kelulusan tepat dan terlambat waktu dilihat dari lama studi mahasiswa.

Variabel Independen

Adapun variabel independen dalam penelitian ini adalah data mahasiswa berupa jenis kelamin, indeks prestasi yang diperoleh dari semester

1 hingga semester 4, beban SKS yang telah ditempuh mahasiswa pada semester iv, lama masa tugas akhir, masa studi mahasiswa, dan jalur masuk. Dimana data-data tersebut merupakan faktor penentu kelulusan mahasiswa tepat atau terlambat waktu, merujuk kepada riset-riset terdahulu diantaranya [14] [15] dan lain sebagainya.

B. Pelaksanaan

Tahapan ini merupakan tahapan aksi dimana proses penelitian dilakukan berdasarkan permasalahan penelitian dan penentuan metode analisisnya. Tahapan ini menguraikan beberapa hal yang dipandang perlu untuk dilakukan, diantaranya pengumpulan data, analisis data, dan pelaporan hasil.

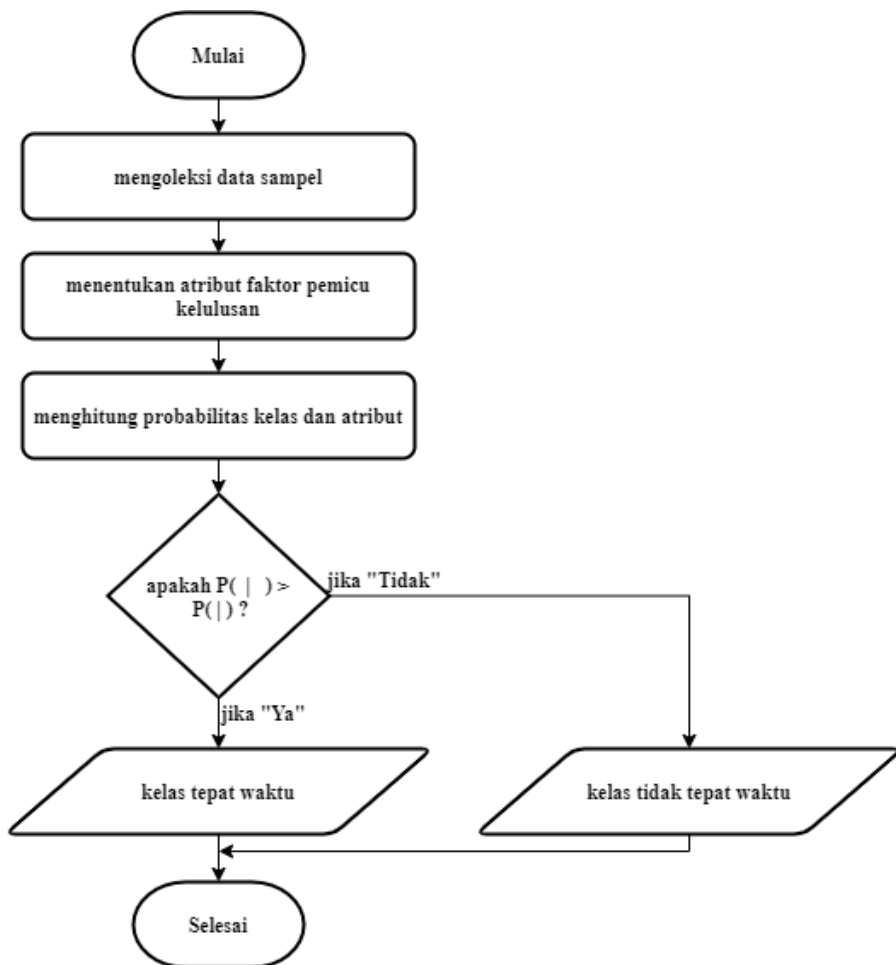
Teknik pengumpulan Data

Penelitian ini menggunakan data sekunder. Dimana data sekunder menurut Sugiono (2008, 402) pada artikel yang dituliskan oleh [16] adalah sumber data yang tidak langsung diberikan kepada pengumpul data. Dengan kata lain, pengumpul data (dalam penelitian / yang memerlukan data sampel), mendapatkan data dari pihak pengumpul data pertama. Biasanya data sampel didapat dari basisdata institusi atau lembaga yang sedang diteliti. Sesuai dengan maksud dari data sekunder itu, pada penelitian ini data sekunder diperoleh dari basisdata Program Studi Ilmu Komputer UIN Sumatera Utara Medan.

Pengolahan atau Analisis Data

Pengolahan data, atau analisis data pada penelitian ini menggunakan metode klasifikasi *naive bayes*. Analisis data penelitian disini bertujuan untuk mengklasifikasikan mahasiswa yang termasuk kedalam kelompok kelas dengan kelulusan tepat waktu, dan kelompok kelas dengan kelulusan terlambat (tidak tepat waktu).

Analisis metode klasifikasi *naive bayes* disini berupa pemodelan matematis yang menghasilkan model algoritma untuk memprediksi mahasiswa yang memungkinkan dapat lulus tepat waktu dan mahasiswa yang memungkinkan lulus terlambat (tidak tepat waktu). Model algoritma prediksi dengan *naive bayes* ini nantinya, dapat diterapkan pula dalam bentuk aplikasi komputer untuk prediksi kelulusan mahasiswa dimasa mendatang. Adapun tahapan metode klasifikasi *naive bayes* pada penelitian ini dapat dilihat pada bagan cara kerja metode di bawah ini :



Gambar 3.2 Bagan Cara Kerja *Naive Bayes*

Penyusunan Laporan

Tahapan ini merupakan tahapan akhir dari penelitian. Dimana laporan berisikan berbagai hal terkait penelitian, mulai dari uraian tentang tahapan persiapan, pelaksanaan, hingga tahapan analisis dan penyajian hasil penelitian, yang akan dibagi dalam bab-bab seperti pendahuluan, tinjauan pustaka, metodologi, serta hasil dan pembahasan, ditutup bab dengan simpulan dan saran sebagai penutup laporan.

Penyusunan laporan perlu dilakukan untuk dokumentasi dan publikasi, sehingga hasil penelitian dapat dijadikan sebagai rujukan bagi penelitian dimasa mendatang. Dengan demikian hasil penelitian benar-benar memberikan kontribusi kepada masyarakat luas, khususnya untuk kalangan peneliti. Publikasi berupa artikel jurnal juga akan dilakukan, yang diupayakan publikasi ke penerbitan jurnal terakreditasi SINTA minimal yang memiliki akreditasi peringkat 3 atau 4, untuk memberikan kontribusi yang banyak kepada masyarakat yang membutuhkan seperti misalnya mahasiswa bidang sains matematika, keteknikan dan informatika, serta kalangan peneliti dan akademisi.

BAB IV

ANALISIS DATA DAN HASIL

Pada bagian ini, analisis dan hasilnya akan disajikan untuk menunjukkan terlaksananya penelitian dengan baik. Lebih lanjut penjelasan analisis data dan pemaparan hasil dapat dilihat pada penjelasan di bawah ini :

A. Analisis Kebutuhan

Untuk memulai penjelasan pada tahapan ini, dilakukan analisis kebutuhan data pelatihan (*training*) dan data pengujian (*testing*). Baik data pelatihan maupun pengujian adalah 44 mahasiswa Program Studi Ilmu Komputer UIN Sumatera Utara Medan angkatan tahun 2015 dan 2016, yang lulus pada 5 waktu sidang *munaqassyah* (seminar hasil) yaitu pada Nopember 2019, Februari 2020, Juli-Agustus 2020, dan Oktober 2020. Adapun data pelatihan dipilih dari 4 waktu sidang sebanyak 38 data lulusan, sementara 6 mahasiswa yang lulus pada Oktober 2020 akan dijadikan sebagai data pengujian.

Tujuan penelitian ini adalah mencari model prediksi kelulusan mahasiswa tepat waktu atau terlambat (tidak tepat waktu), yang dimaksudkan sebagai variabel dependen. Sementara data-data yang bertindak sebagai variabel independennya disajikan pada tabel 4.1, yang merupakan data awal dan belum melalui tahapan proses apapun. Data-datanya diuraikan pada tabel berikut :

Tabel 4.1 Data Asli Kelulusan

NO	JK	IPS 1	IPS 2	IPS 3	IPS 4	SKS 4	LAMA TA	MASA STUDI	JALUR MASUK	KL
1	L	3,29	3,26	3,32	3,24	86	2	4,5	MANDIRI	TL

2	P	3,29	3,17	3,5	3,21	89	4	5	MANDIRI	TL
3	L	3,14	3,35	3,36	3,67	86	3	5	MANDIRI	TL
4	L	3,76	3,84	3,77	3,73	89	4	5	MANDIRI	TL
5	L	3,48	3,09	3,32	3,24	86	2	4,5	MANDIRI	TL
6	L	3,29	3,26	3,18	3,24	86	4	5	MANDIRI	TL
7	P	3,57	3,32	3,45	3,23	89	3	5	MANDIRI	TL
8	P	3,52	3,16	3,41	3,23	89	4	5	MANDIRI	TL
9	P	3,38	3,22	3,55	3,33	89	2	4,5	MANDIRI	TL
10	P	3,38	3,04	3,32	3,24	86	4	5	MANDIRI	TL
11	P	3,43	3,22	3,32	3,33	86	4	5	MANDIRI	TL
12	L	3,14	3,22	3,5	3,46	89	4	5	MANDIRI	TL
13	L	3,57	3,48	3,23	3,5	89	4	5	MANDIRI	TL
14	P	3,71	3,48	3,55	3,41	89	2	4,5	MANDIRI	TL
15	P	3,65	3,12	3,27	3,41	89	4	5	MANDIRI	TL
16	P	3,57	3,24	3,41	3,32	89	2	4,5	MANDIRI	TL
17	P	3,57	3,72	3,55	3,41	89	2	4,5	MANDIRI	TL
18	P	3,75	3,28	3,77	3,5	89	2	4,5	MANDIRI	TL
19	L	3,05	2,96	2,95	3,14	86	4	5	MANDIRI	TL

20	P	3,76	3,68	3,64	3,41	89	2	4,5	MANDIRI	TL
21	P	3,62	3,12	3,32	3,09	89	4	5	MANDIRI	TL
22	L	3,25	3,13	3,23	3,43	86	4	5	MANDIRI	TL
23	P	3,52	3,52	3,77	3,55	89	2	4,5	MANDIRI	TL
24	P	3,76	3,65	3,54	3,38	91	2	4,5	MANDIRI	TL
25	L	3,2	3,17	3,55	3,08	89	4	5	MANDIRI	TL
26	L	3,57	3,52	3,32	3,32	89	3	5	MANDIRI	TL
27	P	3,38	3,83	3,64	3,79	89	2	4,5	MANDIRI	TL
28	P	3,57	3	3,5	3,36	89	2	4,5	MANDIRI	TL
29	L	3,43	3,17	3,23	3,33	86	4	5	MANDIRI	TL
30	L	3,76	3,16	3,41	3,5	89	2	4,5	MANDIRI	TL
31	L	3,52	3,5	3,63	3,61	91	1	4	MANDIRI	TP
32	P	3	3,17	3,23	3,33	86	3	5	MANDIRI	TL
33	L	3,2	3,17	3,09	3,19	86	2	4,5	MANDIRI	TL
34	L	3,3	3,17	3,32	3,19	86	3	5	MANDIRI	TL
35	P	3,29	3,52	3,13	3,29	88	3	5	MANDIRI	TL
36	P	3,48	3,17	3,32	3,14	86	3	5	MANDIRI	TL
37	P	3,29	3,52	3,38	3,48	88	1	4	SBMPTN	TP

38	P	3,82	3,35	3,75	3,54	91	2	4,5	SBMPTN	TL
39	P	3,38	3,43	3,52	3,26	87	2	4,5	MANDIRI	??
40	L	3,57	3,46	3,38	3,45	87	2	4,5	MANDIRI	??
41	P	3,65	3	3,09	2,33	86	5	5	MANDIRI	??
42	L	3,14	3,48	3,04	3,57	88	2	4,5	SBMPTN	??
43	L	3,86	3,52	3,79	3,5	93	2	4,5	MANDIRI	??
44	P	3,05	3,22	3,18	3,33	86	5	5	MANDIRI	??

Untuk memperjelas data pada tabel diatas, berikut ini disampaikan uraian penjelasannya sebagai berikut :

- 1) JK, merupakan singkatan Jenis Kelamin dari setiap mahasiswa yang digunakan dalam data pelatihan dan pengujian.
- 2) IPS, merupakan singkatan Indeks Prestasi Semester, yang dalam data ini diperoleh dari semester 1 sampai semester 4, dan diwakili dengan singkatan IPS 1, IPS 2, IPS 3, dan IPS 4.
- 3) SKS 4, merupakan total jumlah SKS mahasiswa pada semester 4.
- 4) TA, merupakan singkatan dari Tugas Akhi mahasiswa. TA mewakili informasi lamanya masa tugas akhir yang ditempuh mahasiswa, yang dihitung dalam satuan semester.
- 5) Masa Studi, yakni masa studi yang dilalui oleh mahasiswa sampai memperoleh gelar sarjana komputer (S.Kom), yang dihitung dalam satuan tahun.
- 6) Jalur Masuk, yang merupakan atribut yang menerangkan mahasiswa masuk kuliah melalui jalur Seleksi Bersama

Mahasiswa Perguruan Tinggi Negeri (SBMPTN) atau melalui jalur mandiri.

- 7) KL, merupakan singkatan dari Kelulusan yang merupakan kelas yang akan dicari dari model prediksi. Dimana kelulusan tepat waktu akan diwakili dengan kelas TP, dan kelulusan terlambat (tidak tepat waktu) diwakili dengan kelas TL.

Disamping data-data tersebut, terdapat data-data berupa NIM, angkatan (tahun masuk), dan nama. Data-data tersebut tidak termasuk kepada faktor pemicu kelulusan mahasiswa tepat waktu. Dengan demikian data-data tersebut tidak disajikan dalam tabel data sampel. Selanjutnya, data lulusan diatas akan dikelompokkan dalam dua bagian menurut kegunaannya sebagai data pelatihan dan data pengujian yang disajikan pada tabel 4.2 dan 4.3 berikut :

Tabel 4.2 Data Pelatihan

NO	JK	IPS 1	IPS 2	IPS 3	IPS 4	SKS 4	LAMA TA	MASA STUDI	JALUR MASUK	KL
1	L	3,29	3,26	3,32	3,24	86	2	4,5	MANDIRI	TL
2	P	3,29	3,17	3,5	3,21	89	4	5	MANDIRI	TL
3	L	3,14	3,35	3,36	3,67	86	3	5	MANDIRI	TL
4	L	3,76	3,84	3,77	3,73	89	4	5	MANDIRI	TL
5	L	3,48	3,09	3,32	3,24	86	2	4,5	MANDIRI	TL
6	L	3,29	3,26	3,18	3,24	86	4	5	MANDIRI	TL
7	P	3,57	3,32	3,45	3,23	89	3	5	MANDIRI	TL

8	P	3,52	3,16	3,41	3,23	89	4	5	MANDIRI	TL
9	P	3,38	3,22	3,55	3,33	89	2	4,5	MANDIRI	TL
10	P	3,38	3,04	3,32	3,24	86	4	5	MANDIRI	TL
11	P	3,43	3,22	3,32	3,33	86	4	5	MANDIRI	TL
12	L	3,14	3,22	3,5	3,46	89	4	5	MANDIRI	TL
13	L	3,57	3,48	3,23	3,5	89	4	5	MANDIRI	TL
14	P	3,71	3,48	3,55	3,41	89	2	4,5	MANDIRI	TL
15	P	3,65	3,12	3,27	3,41	89	4	5	MANDIRI	TL
16	P	3,57	3,24	3,41	3,32	89	2	4,5	MANDIRI	TL
17	P	3,57	3,72	3,55	3,41	89	2	4,5	MANDIRI	TL
18	P	3,75	3,28	3,77	3,5	89	2	4,5	MANDIRI	TL
19	L	3,05	2,96	2,95	3,14	86	4	5	MANDIRI	TL
20	P	3,76	3,68	3,64	3,41	89	2	4,5	MANDIRI	TL
21	P	3,62	3,12	3,32	3,09	89	4	5	MANDIRI	TL
22	L	3,25	3,13	3,23	3,43	86	4	5	MANDIRI	TL
23	P	3,52	3,52	3,77	3,55	89	2	4,5	MANDIRI	TL
24	P	3,76	3,65	3,54	3,38	91	2	4,5	MANDIRI	TL

25	L	3,2	3,17	3,55	3,08	89	4	5	MANDIRI	TL
26	L	3,57	3,52	3,32	3,32	89	3	5	MANDIRI	TL
27	P	3,38	3,83	3,64	3,79	89	2	4,5	MANDIRI	TL
28	P	3,57	3	3,5	3,36	89	2	4,5	MANDIRI	TL
29	L	3,43	3,17	3,23	3,33	86	4	5	MANDIRI	TL
30	L	3,76	3,16	3,41	3,5	89	2	4,5	MANDIRI	TL
31	L	3,52	3,5	3,63	3,61	91	1	4	MANDIRI	TP
32	P	3	3,17	3,23	3,33	86	3	5	MANDIRI	TL
33	L	3,2	3,17	3,09	3,19	86	2	4,5	MANDIRI	TL
34	L	3,3	3,17	3,32	3,19	86	3	5	MANDIRI	TL
35	P	3,29	3,52	3,13	3,29	88	3	5	MANDIRI	TL
36	P	3,48	3,17	3,32	3,14	86	3	5	MANDIRI	TL
37	P	3,29	3,52	3,38	3,48	88	1	4	SBMPTN	TP
38	P	3,82	3,35	3,75	3,54	91	2	4,5	SBMPTN	TL

Tabel 4.3 Data Pengujian

NO	JK	IPS 1	IPS 2	IPS 3	IPS 4	SKS 4	LAMA TA	MASA STUDI	JALUR MASUK	KL
39	P	3,38	3,43	3,52	3,26	87	2	4,5	MANDIRI	??
40	L	3,57	3,46	3,38	3,45	87	2	4,5	MANDIRI	??

41	P	3,65	3	3,09	2,33	86	5	5	MANDIRI	??
42	L	3,14	3,48	3,04	3,57	88	2	4,5	SBMPTN	??
43	L	3,86	3,52	3,79	3,5	93	2	4,5	MANDIRI	??
44	P	3,05	3,22	3,18	3,33	86	5	5	MANDIRI	??

B. Transformasi Data

Untuk memudahkan proses analisis data pelatihan dan data pengujian perlu ditransformasikan ke dalam bentuk yang sesuai dengan metode analisisnya. Adapun transformasi data pelatihan dan pengujian yang dimaksud dapat dilihat pada tabel-tabel dibawah ini :

Tabel 4.4 Transformasi data pelatihan

J K	IPS1	IPS2	IPS3	IPS4	SKS4	TA	STUD I	JAL UR	K L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L

	=3,00	=3,00	=3,00	=3,00	>=80	>1	I>4		L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA <1	STUD I<=4	M	T P
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	M	T L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA <1	STUD I<=4	S	T P
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STUD I>4	S	T L

Tabel 4.5 Transformasi data pengujian

J K	IPS1	IPS2	IPS3	IPS4	SKS4	TA	STU DI	JAL UR	K L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STU DI>4	M	T L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STU DI>4	M	T L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4<3 ,00	SKS4 >=80	TA >1	STU DI>4	M	T L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STU DI>4	S	T L
L	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STU DI>4	M	T L
P	IPS1> =3,00	IPS2> =3,00	IPS3> =3,00	IPS4> =3,00	SKS4 >=80	TA >1	STU DI>4	M	T L

C. Penyajian Data Analisis

Sebelum analisis data dilakukan, data yang telah dikoleksi akan dipetakan berdasarkan kriteria status kelulusannya kedalam bentuk yang mudah dibaca. Adapun pemetaan datanya disajikan pada tabel-tabel berikut :

Jenis Kelamin

Berikut ini adalah penyajian data kelulusan tepat waktu dan terlambat berdasarkan jenis kelaminnya. Data kelulusan yang dimaksud dapat dilihat berdasarkan tabel berikut :

Tabel 4.6 Kelulusan Berdasarkan Jenis Kelamin

NO	JK	KL		
		TP	TL	JLH L/P
1	L	1	15	16
2	P	1	21	22
JUMLAH		2	36	38

Indeks Prestasi Mahasiswa (IPS)

Dalam penelitian ini, IPS yang digunakan adalah IP semester 1 sampai dengan semester 4, yang disingkat dengan IPS 1, IPS 2, IPS 3, dan IPS 4. IPS yang baik merupakan indikator kemampuan mahasiswa. Yang secara logis memungkinkan mahasiswa mencapai kelulusan tepat waktu, bahkan lebih awal. Sejalan dengan rasionalitas itulah, proses analisis data pada penelitian ini akan merujuk kepada IPS $\geq 3,00$ yang mendukung kelulusan tepat waktu, dan IPS $< 3,00$ mendukung kelulusan terlambat. Penyajian data IPS 1 - IPS 4 disampaikan pada uraian tabel-tabel berikut :

Tabel 4.7 Kelulusan Berdasarkan IP Semester 1

IPS	MAHASISWA	TP	TL
$\geq 3,00$	38	2	36
$< 3,00$	0	0	0
JLH	38	2	36

Tabel 4.8 Kelulusan Berdasarkan IP Semester 2

IPS	MAHASISWA	TP	TL
$\geq 3,00$	37	2	35
$< 3,00$	1	0	1
JLH	38	2	36

Tabel 4.9 Kelulusan Berdasarkan IP Semester 3

IPS	MAHASISWA	TP	TL
$\geq 3,00$	37	2	35
$< 3,00$	1	0	1
JLH	38	2	36

Tabel 4.10 Kelulusan Berdasarkan IP Semester 4

IPS	MAHASISWA	TP	TL
$\geq 3,00$	38	2	36
$< 3,00$	0	0	0
JLH	38	2	36

Jumlah SKS Semester 4

Secara normal mahasiswa yang duduk di tahun kedua atau pada semester 4, telah menyelesaikan beban kuliah 80 sks dari 148 sks yang harus diselesaikan hingga lulus. Sementara pada tabel data yang telah dikumpulkan, rata-rata mahasiswa telah menyelesaikan 86, 88, 89, dan 91 sks. Rasionalnya, semakin cepat beban sks yang telah diselesaikan, akan semakin cepat mencapai kelulusan mengingat beban sks semakin cepat pula

selesai. Untuk itu penelitian ini menerapkan pendekatan bahwa mahasiswa memungkinkan menempuh kelulusan tepat waktu atau bahkan lebih awal bila mahasiswa telah menyelesaikan ≥ 80 sks di semester 4, dan sebaliknya memungkinkan mencapai kelulusan tidak tepat waktu apabila sks di semester tersebut < 80 . Adapun data yang diamati pada penelitian ini mahasiswa telah melampaui beban sks >80 , seperti yang ditampilkan pada tabel berikut :

Tabel 4.11 Kelulusan Berdasarkan Sks Semester 4

SKS	MAHASISWA	TP	TL
≥ 80	38	2	36
< 80	0	0	0
JLH	38	2	36

Lama Tugas Akhir

Tugas akhir yang diterapkan pada Program Studi Ilmu Komputer UIN Sumatera Utara Medan adalah laporan riset skripsi. Umumnya masa tugas akhir mahasiswa beragam. Tentu semakin cepat dan giat mahasiswa dalam menyelesaikan tugas akhir, akan mencapai kelulusan lebih cepat. Adapun masa tugas akhir rata-rata mahasiswa menghabiskan waktu 1, 2, 3, hingga 4 semester. Penelitian ini menggunakan pendekatan bahwa mahasiswa memungkinkan lulus tepat waktu atau lebih awal bila masa tugas akhir dapat diselesaikan tidak kurang dari 1 semester atau < 6 bulan, dan untuk mahasiswa yang menghabiskan waktu > 1 semester atau > 6 bulan, mahasiswa mungkin mencapai kelulusan tidak tepat waktu.

Tabel 4.12 Kelulusan Berdasarkan Lama Tugas Akhir

NO	LAMA TUGAS AKHIR	KL	
		TP	TL
1	≤ 1	2	0
2	> 1	0	36
JLH		2	36

Masa Studi

Masa studi mahasiswa secara normal adalah 4 tahun atau 8 semester, yang juga dapat ditempuh lebih cepat dari tahun itu. Indikator ketepatan waktu dan terlambatnya kelulusan mahasiswa dilihat dari masa studi ini, yaitu mahasiswa lulus tepat waktu bilamana mahasiswa lulus ≤ 4 tahun, dan lulus terlambat bilamana mahasiswa lulus > 4 tahun. Adapun data masa studi untuk lulusan mahasiswa yang sedang dijadikan data penelitian ini adalah sesuai pada uraian tabel berikut :

Tabel 4.13 Kelulusan Berdasarkan Masa Studi

NO	MASA STUDI	KL	
		TP	TL
1	≤ 4	2	0
2	> 4	0	36
JLH		0	36

Jalur Masuk

Dari mekanisme yang ada, mahasiswa dapat ke UIN Sumatera Utara Medan melalui beberapa jalur masuk SNMPT, SPAN PTKIN, SBMPTN, UMPTKIN, dan jalur MANDIRI. Namun, untuk mahasiswa angkatan pertama (2015), jalur masuk kuliah masih melalui jalur mandiri. Adapun data

mahasiswa kelulusan berdasarkan jalur seleksi masuk kuliah dijelaskan pada tabel 4.14 sebagai berikut :

Tabel 4.14 Kelulusan Berdasarkan Jalur Masuk Kuliah

NO	JALUR MASUK	KL	
		TP	TL
1	SBMPTN	1	1
2	MANDIRI	1	35
JUMLAH		2	36

D. Analisis dan Pemodelan *Naïve Bayes*

Ini adalah tahap ini, dimana tahapan analisis terhadap data dilakukan dengan membuat model *naïve bayes* dalam melakukan prediksi kelulusan tepat waktu dan terlambat untuk dapat diterapkan kepada mahasiswa dimasa mendatang. Tahapan-tahapan analisis dengan metode *naïve bayes*-nya diuraikan sebagai berikut :

Menghitung Probabilitas Kelas

Untuk melihat kelayakan suatu metode *naïve bayes*, probabilitas kelayakan setiap kelasnya harus ditentukan terlebih dahulu. Dalam penelitian ini kelas yang dicari adalah kelas kelulusan “tepat waktu” dan kelas kelulusan “tidak tepat waktu atau terlambat”. Sebelum menerapkan teori *naïve bayes* dicari terlebih dahulu probabilitas dari kedua kelas tersebut dengan persamaan 2.1.

Jika melihat data kelulusan pada tabel kelulusan, ada 2 mahasiswa dari 38 mahasiswa keluar lulus. Kelulusan disini diwakili dengan variabel Y, maka probabilitas kelas diperoleh sebagai berikut :

$$P(Y = \text{“Tepat Waktu”}) = 2/38 = 0,053$$

$$P(Y = \text{"Tidak Tepat Waktu"}) = 36/38 = 0,947$$

Menghitung Probabilitas Kejadian

Kejadian yang dimaksud disini adalah fakta dari faktor-faktor yang memengaruhi kelulusan tepat dan tidak tepat waktu, dengan menggunakan persamaan yang sama dengan persamaan untuk probabilitas kelas yang dicari. Kejadian disini diwakili dengan variabel X, sehingga didapat sebagai berikut :

- 1) Probabilitas Kejadian Atribut X untuk Kelas Tepat Waktu

Untuk X_1 adalah Jenis Kelamin, probabilitasnya adalah :

$$P(X = \text{"L"} \mid Y = \text{"Tepat Waktu"}) = 1/2 = 0,5$$

$$P(X = \text{"P"} \mid Y = \text{"Tepat Waktu"}) = 1/2 = 0,5$$

Untuk X_2 adalah IP Semester 1, probabilitasnya adalah :

$$P(X = \text{"IPS1 } \geq 3,00" \mid Y = \text{"Tepat Waktu"}) = 2/2 = 1$$

$$P(X = \text{"IPS1 } < 3,00" \mid Y = \text{"Tepat Waktu"}) = 0/2 = 0$$

Untuk X_3 adalah IP Semester 2, probabilitasnya adalah :

$$P(X = \text{"IPS2 } \geq 3,00" \mid Y = \text{"Tepat Waktu"}) = 2/2 = 1$$

$$P(X = \text{"IPS2 } < 3,00" \mid Y = \text{"Tepat Waktu"}) = 0/2 = 0$$

Untuk X_4 adalah IP Semester 3, probabilitasnya adalah :

$$P(X = \text{"IPS3 } \geq 3,00" \mid Y = \text{"Tepat Waktu"}) = 2/2 = 1$$

$$P(X = \text{"IPS3 } < 3,00" \mid Y = \text{"Tepat Waktu"}) = 0/2 = 0$$

Untuk X_5 adalah IP Semester 4, probabilitasnya adalah :

$$P(X = \text{"IPS4 } \geq 3,00" \mid Y = \text{"Tepat Waktu"}) = 2/2 = 1$$

$$P(X = \text{"IPS4 } < 3,00" \mid Y = \text{"Tepat Waktu"}) = 0/2 = 0$$

Untuk X_6 adalah Jumlah SKS Semester 4, probabilitasnya adalah :

$$P(X = \text{"SKS4 } \geq 80" \mid Y = \text{"Tepat Waktu"}) = 2/2 = 1$$

$$P(X = \text{"SKS4 } < 80" \mid Y = \text{"Tepat Waktu"}) = 0/2 = 0$$

Untuk X_7 adalah Lama Tugas Akhir, probabilitasnya adalah :

$$P(X = \text{"TA"} \leq 1 \mid Y = \text{"Tepat Waktu"}) = 2/2 = 1$$

$$P(X = \text{"TA"} > 1 \mid Y = \text{"Tepat Waktu"}) = 0/2 = 0$$

Untuk X_8 adalah Masa Studi, yang probabilitasnya adalah :

$$P(X = \text{"STUDI"} \leq 4 \mid Y = \text{"Tepat Waktu"}) = 2/2 = 1$$

$$P(X = \text{"STUDI"} > 4 \mid Y = \text{"Tepat Waktu"}) = 0/2 = 0$$

Untuk X_9 adalah Jalur Masuk PTN, dan probabilitasnya adalah :

$$P(X = \text{"JALUR=SBMPTN"} \mid Y = \text{"Tepat Waktu"}) = 1/2 = 0,5$$

$$P(X = \text{"JALUR=MANDIRI"} \mid Y = \text{"Tepat Waktu"}) = 1/2 = 0,5$$

2) Probabilitas Kejadian Atribut X untuk Kelas Tidak Tepat Waktu

Untuk X_1 adalah Jenis Kelamin, probabilitasnya adalah :

$$P(X = \text{"L"} \mid Y = \text{"Tidak Tepat Waktu"}) = 15/36 = 0,417$$

$$P(X = \text{"P"} \mid Y = \text{"Tidak Tepat Waktu"}) = 21/36 = 0,583$$

Untuk X_2 adalah IP Semester 1, probabilitasnya adalah :

$$P(X = \text{"IPS1"} \geq 3,00 \mid Y = \text{"Tidak Tepat Waktu"}) = 36/36 = 1$$

$$P(X = \text{"IPS1"} < 3,00 \mid Y = \text{"Tepat Waktu"}) = 0/36 = 0$$

Untuk X_3 adalah IP Semester 2, probabilitasnya adalah :

$$P(X = \text{"IPS2"} \geq 3,00 \mid Y = \text{"Tidak Tepat Waktu"}) = 35/36 = 0,972$$

$$P(X = \text{"IPS2"} < 3,00 \mid Y = \text{"Tidak Tepat Waktu"}) = 1/36 = 0,027$$

Untuk X_4 adalah IP Semester 3, probabilitasnya adalah :

$$P(X = \text{"IPS3"} \geq 3,00 \mid Y = \text{"Tidak Tepat Waktu"}) = 35/36 = 0,972$$

$$P(X = \text{"IPS3"} < 3,00 \mid Y = \text{"Tidak Tepat Waktu"}) = 1/36 = 0,027$$

Untuk X_5 adalah IP Semester 4, probabilitasnya adalah :

$$P(X = \text{"IPS4"} \geq 3,00 \mid Y = \text{"Tepat Waktu"}) = 36/36 = 1$$

$$P(X = \text{"IPS4"} < 3,00 \mid Y = \text{"Tepat Waktu"}) = 0/36 = 0$$

Untuk X_6 adalah Jumlah SKS Semester 4, probabilitasnya adalah :

$$P(X = \text{"SKS4"} \geq 80 \mid Y = \text{"Tidak Tepat Waktu"}) = 36/36 = 1$$

$$P(X = \text{"SKS4"} < 80 \mid Y = \text{"Tidak Tepat Waktu"}) = 0/36 = 0$$

Untuk X_7 adalah Lama Tugas Akhir, probabilitasnya adalah :

$$P(X = \text{"TA"} \leq 1 \mid Y = \text{"Tidak Tepat Waktu"}) = 0/36 = 0$$

$$P(X = \text{"TA"} > 1 \mid Y = \text{"Tidak Tepat Waktu"}) = 36/36 = 1$$

Untuk X_8 adalah Masa Studi, yang probabilitasnya adalah :

$$P(X = \text{"STUDI"} \leq 4 \mid Y = \text{"Tidak Tepat Waktu"}) = 0/36 = 0$$

$$P(X = \text{"STUDI"} > 4 \mid Y = \text{"Tidak Tepat Waktu"}) = 36/36 = 1$$

Untuk X_9 adalah Jalur Masuk PTN, dan probabilitasnya adalah :

$$P(X = \text{"JALUR=SBMPTN"} \mid Y = \text{"Tidak Tepat Waktu"}) = 1/36 = 0,027$$

$$P(X = \text{"JALUR=MANDIRI"} \mid Y = \text{"Tidak Tepat Waktu"}) = 35/36 = 0,972$$

3) Menghitung probabilitas label "Tepat Waktu"

Setelah menghitung probabilitas semua kejadian yang memiliki keluaran Tepat Waktu dari data pelatihan diatas, maka selanjutnya akan di tentukan probabilitas dari label "Tepat Waktu" sebagai berikut :

$$P(Y = \text{"Tepat Waktu"}) = P(X_1(P) \mid Y(\text{"Tepat Waktu"})) *$$

$$P(X_2(>=3) \mid Y(\text{"Tepat Waktu"})) * P(X_3(>=3) \mid Y(\text{"Tepat Waktu"})) *$$

$$P(X_4(>=3) \mid Y(\text{"Tepat Waktu"})) * P(X_5(>=3) \mid Y(\text{"Tepat Waktu"})) *$$

$$P(X_6(>=80) \mid Y(\text{"Tepat Waktu"})) * P(X_7(\leq 1) \mid Y(\text{"Tepat Waktu"})) *$$

$$P(X_8(\leq 4) \mid Y(\text{"Tepat Waktu"})) * P(X_9(\text{SBMPTN}) \mid Y(\text{"Tepat Waktu"}))$$

$$* P(Y \mid \text{"Tepat Waktu"})$$

$$P(Y = \text{"Tepat Waktu"}) = 0,053 * (0,5 * 1 * 1 * 1 * 1 * 1 * 1 * 1 * 1 * 0,5)$$

$$= 0,01325$$

4) Menghitung probabilitas label "Tidak Tepat Waktu"

$$P(Y = \text{"Tidak Tepat Waktu"}) = P(X_1(P) \mid Y(\text{"Tidak Tepat Waktu"})) *$$

$$P(X_2(>=3) \mid Y(\text{"Tidak Tepat Waktu"})) * P(X_3(>=3) \mid Y(\text{"Tidak Tepat Waktu"})) *$$

$$P(X_4(>=3) \mid Y(\text{"Tidak Tepat Waktu"})) *$$

$$\begin{aligned}
& P(X_5(\geq 3)|Y(\text{"Tidak Tepat Waktu"})) * P(X_6(\geq 80)|Y(\text{"Tidak Tepat Waktu"})) * P(X_7(>1)|Y(\text{"Tidak Tepat Waktu"})) * P(X_8(>4)|Y(\text{"Tidak Tepat Waktu"})) * P(X_9(\text{MANDIRI})|Y(\text{"Tidak Tepat Waktu"})) * P(Y|\text{"Tidak Tepat Waktu"}) \\
& P(Y=\text{"Tidak Tepat Waktu"}) = 0,947 * (0,583 * 1 * 0,972 * 0,972 * 1 * 1 * 1 * 1 * 0,972) = 0,50701
\end{aligned}$$

Membandingkan Probabilitas Label atau Kelas

Setelah melihat uraian perhitungan dari semua atribut diatas (yaitu X_1 - X_9 atau probabilitas JK, IPS1, IPS2, IPS3, IPS4, SKS4, TA, STUDI, dan JALUR, yang diakhiri dengan mengakumulasi nilai-nilai probabilitas atribut tersebut untuk probabilitas kelas yang dicari. Adapun probabilitas kelas yang dicari adalah kelas kelulusan tepat waktu dan kelas kelulusan tidak tepat waktu. Maka selanjutnya membandingkan nilai probabilitas dari kedua kelas tersebut.

Jika dilihat dari akumulasi nilai probabilitas kelas tersebut, yang masing-masing yaitu 0,01325 untuk kelas tepat waktu, dan 0,50701 untuk kelas tidak tepat waktu. Untuk itu perbandingan kedua kelas adalah bahwa :

$$P(Y = \text{"Tepat Waktu"}) < P(Y = \text{"Tidak Tepat Waktu"})$$

Atau

$$P(Y = \text{"Tidak Tepat Waktu"}) > P(Y = \text{"Tepat Waktu"}).$$

Dengan demikian, didapatkan kesimpulan bahwa dari data pelatihan diatas, mahasiswa dominan masuk ke dalam kategori kelas "Tidak Tepat Waktu" dengan rinciang 36 mahasiswa keluar lulus tidak tepat waktu, dan 2 mahasiswa keluar lulus tepat waktu.

Dengan model prediksi kelulusan menggunakan metode klasifikasi *naïve bayes* diatas, jika diterapkan pada data pengujian sebanyak 6

mahasiswa yang keluar lulus pada sidang *munaqassah* bulan Oktober 2020, dapat disimpulkan bahwa ke-6 mahasiswa tersebut juga masuk dalam kategori kelas “tidak tepat waktu.”

E. Penerapan Model Dengan RapidMiner

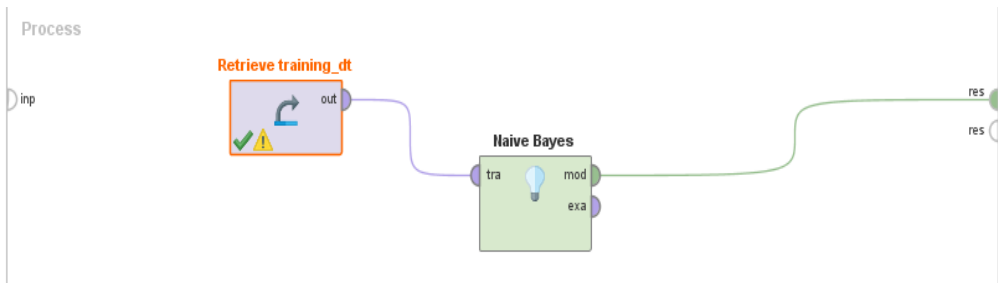
Kelebihan RapidMiner dalam melakukan analisis data diskrit dan numerik, membuat aplikasi ini populer dikalangan peneliti. Kelebihannya juga didukung oleh fungsi-fungsi dari metode-metode analisis data serta kemudahannya dalam menggunakannya. Model prediksi kelulusan tepat waktu dan tidak tepat waktu mahasiswa dalam penelitian ini menggunakan RapidMiner versi 7.6.000 yang dioperasikan pada perangkat laptop processor Amd A9-9420 Radeon R5, 5 Compute Cores 2c+3g 3.00 Ghz, dengan memori 4 GB serta media penyimpanan 1 Terra Byte.

Dengan kebutuhan perangkat untuk operasi RapidMiner 7.6.00 seperti yang diuraikan diatas, akan disajikan dibawah ini hasil analisis data dan model prediksinya dengan metode klasifikasi *naïve bayes* berikut ini :

Model Prediksi dengan *Naïve Bayes*

Untuk menerapkan model prediksi kelulusan dengan RapidMiner dibutuhkan data pelatihan untuk tahapan pelatihan data dan data pengujian untuk pengujian model prediksi. Data pelatihan ini, diambil dari 38 data lulusan yang ditetapkan sebagai data pelatihan, dan 6 data digunakan untuk pengujian. Baik data pelatihan maupun data pengujian ini, telah ditransformasikan kedalam bentuk data yang sesuai seperti yang ditampilkan pada tabel 4.4 dan 4.5.

Tahap pertama data pelatihan dan pengujian diimpor terlebih dahulu kedalam aplikasi RapidMiner. Selanjutnya data pelatihan ditambahkan ke halaman antarmuka *Process*, yang dalam penerapan modelnya disusun seperti pada antarmuka berikut :



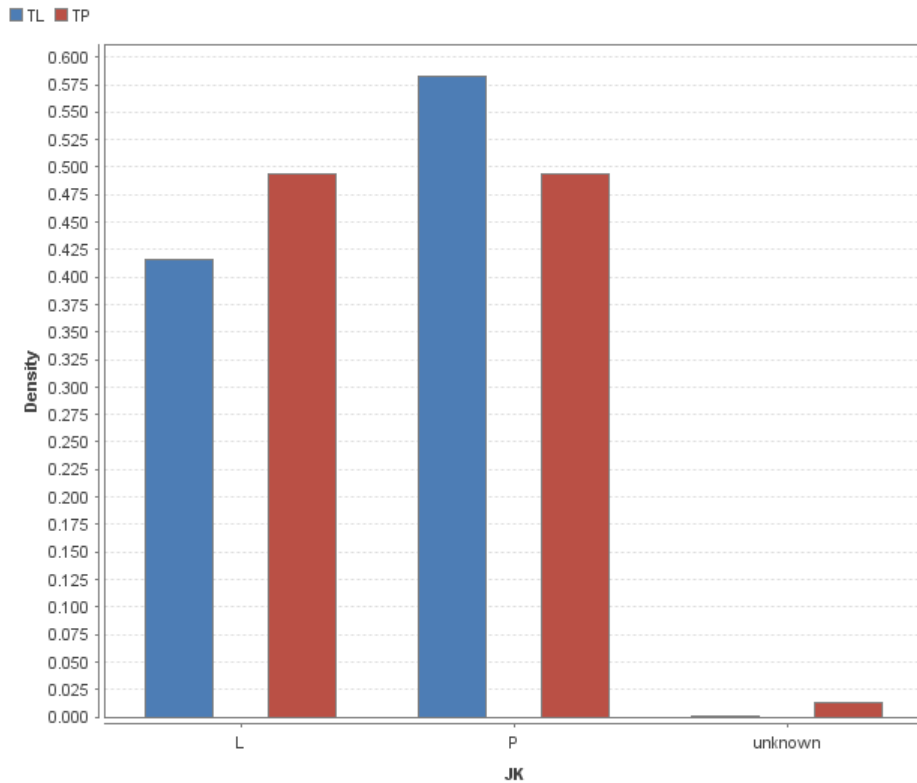
Gambar 4.1 Antarmuka palatihan model prediksi dengan *naïve bayes*

Selanjutnya proses pelatihan data dilakukan dengan metode *naïve bayes*, dan ini akan menghasilkan model prediksi kelulusan mahasiswa tepat waktu dan tidak tepat waktu. Dimana model prediksi ini diporelah berdasarkan nilai probabilitas masing-masing atribut dari proses pelatihan data dengan rincian nilai probabilitasnya disajikan pada keluaran berikut :

Attribute	Parameter	TL	TP
JK	value=L	0.416	0.494
JK	value=P	0.583	0.494
JK	value=unknown	0.001	0.013
IPS1	value=IPS1>=3,00	0.999	0.987
IPS1	value=unknown	0.001	0.013
IPS2	value=IPS2>=3,00	0.971	0.975
IPS2	value=IPS2<3,00	0.028	0.013
IPS2	value=unknown	0.001	0.013
IPS3	value=IPS3>=3,00	0.971	0.975
IPS3	value=IPS3<3,00	0.028	0.013
IPS3	value=unknown	0.001	0.013
IPS4	value=IPS4>=3,00	0.999	0.987
IPS4	value=unknown	0.001	0.013
SKS4	value=SKS4>=80	0.999	0.987
SKS4	value=unknown	0.001	0.013
TA	value=TA>1	0.999	0.013
TA	value=TA<1	0.001	0.975
TA	value=unknown	0.001	0.013
STUDI	value=STUDI>4	0.999	0.013
STUDI	value=STUDI<=4	0.001	0.975
STUDI	value=unknown	0.001	0.013
JALUR	value=M	0.971	0.494
JALUR	value=S	0.028	0.494
JALUR	value=unknown	0.001	0.013

Gambar 4.2 Nilai probabilitas atribut dari pelatihan dengan rapidminer

Sementara sebaran data prediksi kelulusan tepat waktu (TP), dan kelulusan tidak tepat waktu (TL) berdasarkan jenis kelamin dalam pelatihan dengan RapiMiner disajikan pada grafik berikut :

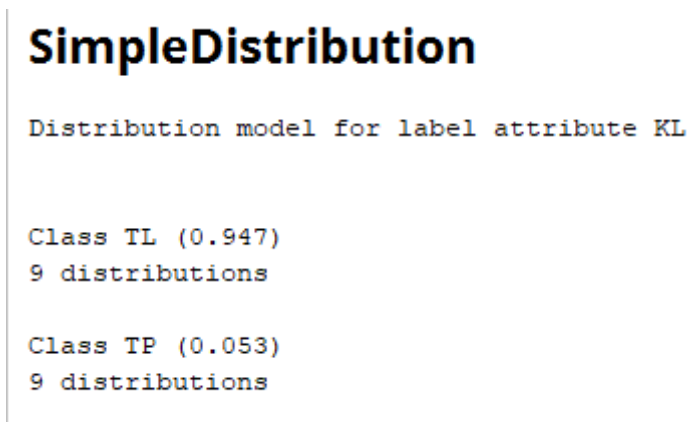


Gambar 4.3 Grafik sebaran data lulusan berdasarkan jenis kelamin

Hasil pelatihan data dengan RapidMiner juga menampilkan distribusi statistik dari setiap atribut kelas yang digunakan. Secara umum tabel distribusi statistiknya ditampilkan pada gambar 4.4 dan gambar 4.5.

Name	Type	Missing	Statistics		
JK	Binominal	0	Least L (16)	Most P (22)	Values P (22), L (16)
IPS1	Binominal	0	Least IPS1>=3,00 (38)	Most IPS1>=3,00 (38)	Values IPS1>=3,00 (38)
IPS2	Binominal	0	Least IPS2<3,00 (1)	Most IPS2>=3,00 (37)	Values IPS2>=3,00 (37), IPS2<3,00 (1)
IPS3	Binominal	0	Least IPS3<3,00 (1)	Most IPS3>=3,00 (37)	Values IPS3>=3,00 (37), IPS3<3,00 (1)
IPS4	Binominal	0	Least IPS4>=3,00 (38)	Most IPS4>=3,00 (38)	Values IPS4>=3,00 (38)
SKS4	Binominal	0	Least SKS4>=80 (38)	Most SKS4>=80 (38)	Values SKS4>=80 (38)
TA	Binominal	0	Least TA<1 (2)	Most TA>1 (36)	Values TA>1 (36), TA<1 (2)
STUDI	Binominal	0	Least STUDI<=4 (2)	Most STUDI>4 (36)	Values STUDI>4 (36), STUDI<=4 (2)
JALUR	Binominal	0	Least S (2)	Most M (36)	Values M (36), S (2)
KL	Binominal	0	Least TP (2)	Most TL (36)	Values TL (36), TP (2)

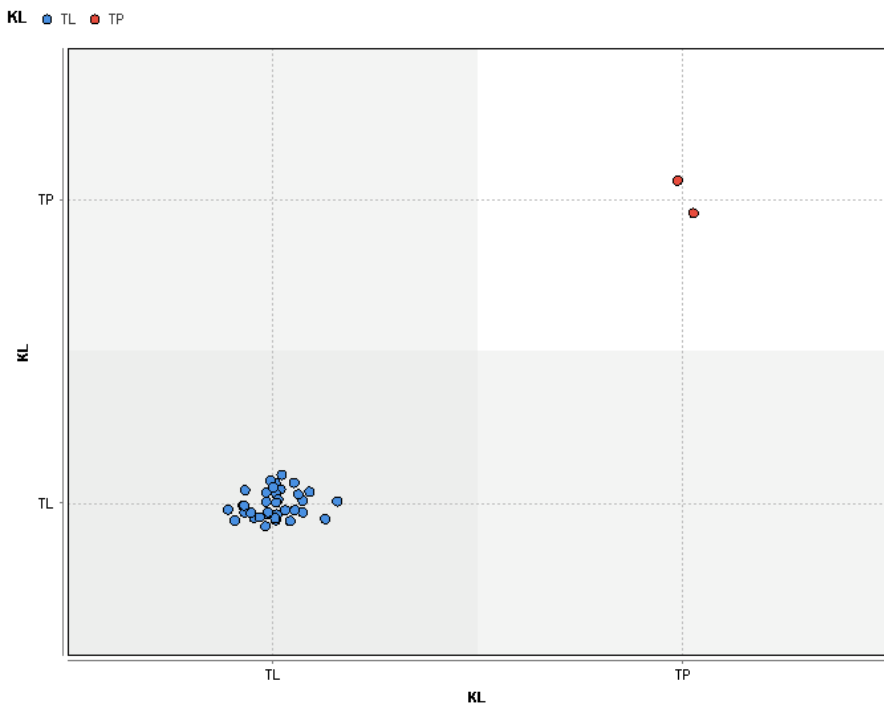
Gambar 4.4 Distribusi statistik dari atribut kelas



Gambar 4.5 Distribusi atribut pelatihan

Dalam *machine learning*, produk aplikasi yang ditawarkan adalah berupa pengelompokan (*clustering*), pengenalan pola (*pattern recognition*), dan peramalan atau prediksi. Aplikasi klasifikasi digunakan untuk tahapan

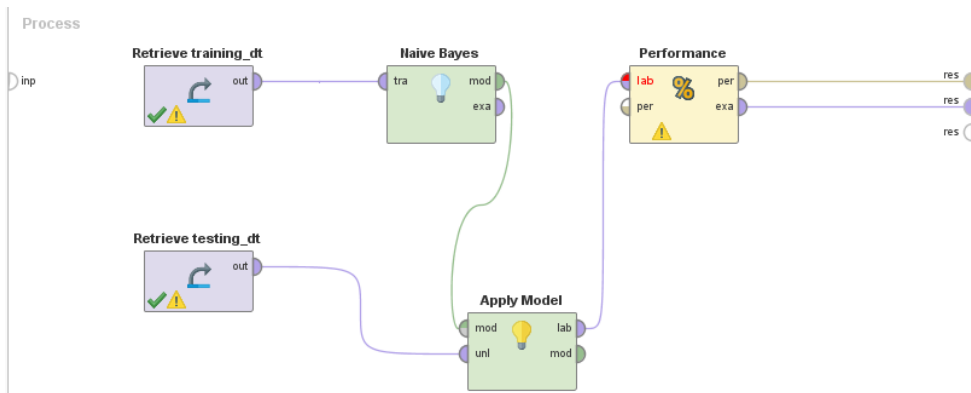
peramalan atau prediksi berdasarkan klasifikasi kelas. Tahap pelatihan data dengan RapidMiner dalam penelitian ini menghasilkan suatu bentuk klasifikasi kelulusan mahasiswa tepat waktu dan tidak tepat waktu, yang diwakili oleh simbol biru dan merah dengan label TP dan TL seperti yang ditampilkan gambar berikut :



Gambar 4.6 Grafik kelas pelatihan data

Pengujian Data dengan Model

Setelah rancangan model dibentuk dan diterapkan dalam melakukan pelatihan data, selanjutnya pengujian data sudah dapat digunakan. Namun dalam proses pembelajaran mesin dengan RapidMiner, perlu mengkonfigurasi data dengan model yang digunakan. Adapun konfigurasi data dan model pelatihan akan tampak seperti pada antarmuka rancangan model berikut ini :



Gambar 4.7 Antarmuka model untuk pengujian data

Setelah rancangan model dikonfigurasi selanjutnya pengujian data dapat dilakukan. Adapun pengujian data menggunakan data lulusan Oktober 2020, yakni sebagai 6 mahasiswa. Setelah dilakukan pelatihan, diperoleh data prediksi kelulusan tepat dan tidak tepat waktu seperti yang ditampilkan pada gambar berikut :

Row...	KL	prediction(KL)	confidence(TL)	confidence(TP)	JK	IPS1
1	TL	TL	1	0	P	IPS1>=3,00
2	TL	TL	1	0	L	IPS1>=3,00
3	TL	?	?	?	P	IPS1>=3,00
4	TL	TL	1	0	L	IPS1>=3,00
5	TL	TL	1	0	L	IPS1>=3,00
6	TL	TL	1	0	P	IPS1>=3,00

IPS2	IPS3	IPS4	SKS4	TA	STUDI	JALUR
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4<3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	S
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M

Gambar 4.8 Hasil prediksi dari pengujian data

Jika dilihat dari gambar 4.8 diatas tampak bahwa model prediksi berhasil melakukan klasifikasi kelulusan. Dari gambar tersebut di peroleh kesimpulan bahwa 5 mahasiswa masuk dalam kategori kelulusan “tidak tepat waktu” serta 1 mahasiswa yang tidak dapat dibaca.

Akurasi Model

Untuk melihat hubungan antara model prediksi dengan pendekatan matematis dan penerapan aplikasi RapidMiner, perlu mempertimbangkan tingkat akurasi model yang digunakan. Akurasi pelatihan model untuk pengujian data yang dilakukan dengan RapidMiner ditampilkan pada gambar *confusion matrix* dan *performace vektor* dengan tingkat akurasi mencapai 100%.

accuracy: 100.00%

	true TL	true TP	class precision
pred. TL	5	0	100.00%
pred. TP	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 4.9 Akurasi Pelatihan Model

PerformanceVector

```
PerformanceVector:  
accuracy: 100.00%  
ConfusionMatrix:  
True:   TL      TP  
TL:     5       0  
TP:     0       0  
kappa:  unknown  
ConfusionMatrix:  
True:   TL      TP  
TL:     5       0  
TP:     0       0
```

Gambar 4.10 *Performance vector*

Jika melihat tabel prediksi pada gambar 4.8, 4.9, dan 4.10 di atas masih menunjukkan performa model yang kurang sempurna karena masih terdapat data yang tidak dapat diklasifikasi, meskipun akurasi yang dihasilkan sudah mencapai 100%. Terkait dengan itu, uji coba dilakukan kembali terhadap data pengujian dan menghasilkan prediksi, performa serta akurasi yang sempurna seperti yang ditampilkan pada gambar 4.11, 4.12 dan 4.13. Dari ketiga gambar tersebut tampak bahwa hasil pengujian terhadap 6 data lulusan menunjukkan klasifikasi kelulusan tidak tepat waktu sesuai dengan model yang dirancang. Sementara akurasi untuk pengujian kedua tersebut menunjukkan akurasi yang sempurna 100%.

ExampleSet (6 examples, 4 special attributes, 9 regular attributes)

Row...	KL	prediction(KL)	confidence(TL)	confidence(TP)	JK	IPS1
1	TL	TL	1.000	0.000	P	IPS1>=3,00
2	TL	TL	1.000	0.000	L	IPS1>=3,00
3	TL	TL	1.000	0.000	P	IPS1>=3,00
4	TL	TL	1.000	0.000	L	IPS1>=3,00
5	TL	TL	1.000	0.000	L	IPS1>=3,00
6	TL	TL	1.000	0.000	P	IPS1>=3,00

IPS2	IPS3	IPS4	SKS4	TA	STUDI	JALUR
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4<3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	S
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M
IPS2>=3,00	IPS3>=3,00	IPS4>=3,00	SKS4>=80	TA>1	STUDI>4	M

Gambar 4.11 Hasil prediksi dari pengujian kedua

accuracy: 100.00%

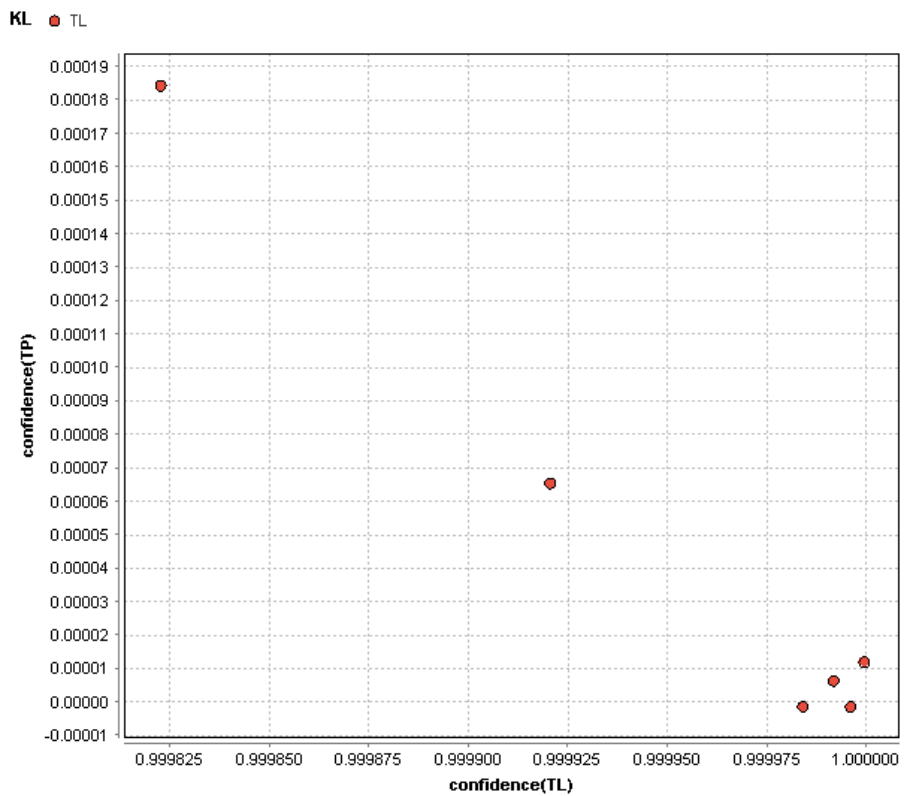
	true TL	true TP	class precision
pred. TL	6	0	100.00%
pred. TP	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 4.12 Akurasi model NB dari pengujian kedua

PerformanceVector

```
PerformanceVector:  
accuracy: 100.00%  
ConfusionMatrix:  
True:   TL   TP  
TL:     6   0  
TP:     0   0  
kappa: unknown  
ConfusionMatrix:  
True:   TL   TP  
TL:     6   0  
TP:     0   0
```

Gambar 4.13 *Performance vector*



Gambar 4.14 Grafik klasifikasi pada ujicoba kedua

Temuan Penelitian

Jika melihat hasil dari model prediksi, baik yang melalui dengan pendekatan matematis metode klasifikasi *naïve bayes*, serta penerapan model prediksi *naïve bayes* dengan RapidMiner diatas, dapat ditarik kesimpulan bahwa keduanya menunjukkan sinkronisasi yang kuat. Dengan kata lain pendekatan matematis metode klasifikasi *naïve bayes* dapat diterapkan untuk memprediksi kelulusan tepat waktu dan tidak tepat waktu mahasiswa. Namun perlu dipertimbangkan pemilihan atribut lain disamping 9 atribut yang telah digunakan. Hal ini dipandang penting mengingat hampir semua kriteria yang dimiliki oleh mahasiswa secara logis memenuhi syarat untuk terpenuhinya kelulusan tepat waktu bahkan lebih awal.

Suatu contoh seperti atribut IP Semester 1 sampai IP Semester 4 yang didominasi oleh lulusan yang memiliki nilai yang baik bahkan beberapa diantaranya mendapat nilai *cum laud*, sementara jumlah beban sks di semester 4 yang diperoleh mahasiswa juga dipandang cukup tinggi, karena telah melampaui batas normal sks mahasiswa di semester 4. Namun, fakta dilapangan mahasiswa dengan kriteria yang tersebut diatas, masuk kedalam klasifikasi kelulusan yang “tidak tepat waktu.” Artinya, ada parameter lain yang juga memengaruhi keterlambatan kelulusan mahasiswa Program Studi Ilmu Komputer UIN Sumatera Utara Medan, yang tidak digunakan dalam penelitian ini. Temuan ini dapat menjadi rujukan bagi kalangan peneliti di masa mendatang untuk melakukan penelitian lanjutan dengan menerapkan atribut-atribut lain, dengan juga memperhatikan fenomena yang terjadi atau sedang berlaku.

BAB V

PENUTUP

A. Kesimpulan

Setelah melakukan tahap demi tahapan penelitian, yang merujuk kepada permasalahan serta rumusan penelitian diatas, hingga kemudian penyajian model dan telaah data pada tahapan analisis serta penerapan RapidMiner yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut :

- 1) Klasifikasi terhadap mahasiswa yang lulus tepat waktu dan tidak tepat waktu, dapat dipecahkan dengan metode *naïve bayes* dan dapat dijadikan model prediksi kelulusan mahasiswa.
- 2) Klasifikasi kelas lulus tepat waktu dan tidak tepat waktu, antara model matematika *naïve bayes* dengan penerapan *tool* RapidMiner memperoleh nilai probabilitas masing-masing atribut dan kelas yang hampir sama.
- 3) Dimana klasifikasi kelas kelulusan untuk model prediksi yang digunakan menunjukkan tingkat akurasi yang tinggi, yakni 100% dengan hasil pelatihan dan pengujian yang tanpa *error*. Ini dibuktikan pada data pengujian model melalui RapidMiner yang menampilkan kelulusan tidak tepat waktu untuk semua *record* data uji, yakni sebanyak 6 data mahasiswa.

B. Saran

Untuk menyempurnakan penelitian ini dimasa mendatang, beberapa saran penulis diantaranya adalah :

- 1) Dengan melihat pencapaian mahasiswa, berupa tingginya indeks prestasi mahasiswa pada semester 1 hingga semester 4, serta jumlah beban sks semester 4 yang cukup tinggi dari normal, perlu dipertimbangkan atribut lain, seperti performa, kepribadian dan

pemahaman mahasiswa terhadap kompetensi bidang program studi, serta sistem dan teknik evaluasi yang digunakan.

- 2) Menambahkan ragam jalur masuk perguruan tinggi, mengingat pada penelitian ini hanya menggunakan 2 jalur saja, yakni jalur SBMPTN dan jalur MANDIRI. Ini dipandang perlu untuk melihat temuan-temuan penelitian lainnya.
- 3) Pemilihan atribut yang digunakan mungkin saja memiliki hubungan yang kuat untuk memicu kelulusan tepat dan tidak tepat waktu, karenanya, pada penelitian selanjutnya perlu menerapkan metode untuk mengukur atribut yang sangat cocok untuk pemodelan prediksi.

DAFTAR REFERENSI

- [1] A. Zed and A. L. I. Muhammad, “Penerapan metode *naive bayes* untuk prediksi menentukan karyawan tetap pada pt. ysp industries indonesia,” STT Pelita Bangsa, Bekasi, 2018.
- [2] N. Amalia, Shaufiah, S. Sa'adah, “Penerapan Teknik Data Mining Untuk Klasifikasi Ketepatan Waktu Lulus Mahasiswa Teknik Informatika Universitas Telkom Menggunakan Algoritma Naive Bayes Classifier”, 2014.
- [3] Abhishek Vijayvargia, "*Machine Learning with Python An Approach to Applied Machine Learning*", INDIA, BPB Publications, 2018
- [4] J. Hurwitz, D. Kirsch, "*Machine Learning For Dummies, IBM Limited Edition*", John Wiley and Sons, Inc., USA, 2018.
- [5] S. Agarwal, "*Data mining: Data mining concepts and techniques*", ICMIRA, 2014.
- [6] Fitri Hidayanti, “Probabilitas dan Statistik : Aplikasi Probabilitas”, L. P. Unas, Jakarta, 2017.
- [7] N. R. I. Husnul, E. R. Prasetya, P. Sadewa, Ajimat, L. I. Purnomo, "Statistik Deskriptif", UNPAM Press, Bekasi, 2019 .
- [8] Tutorialspoint, “*Machine Learning with Python*,” *Tutorials Point Pvt. Ltd.*, India, 2019
- [9] Aprilla Dennis, “Belajar Data Mining dengan RapidMiner,” *Innov. Knowl. Manag. Bus. Glob. Theory Pract. Vols 1 2*, vol. 5, no. 4, pp. 1–5, 2013, [Online].
- [10] J. Letkiewicz, H. Lim, S. Heckman, S. Bartholomae, J. Fox, and C. Montalto, “The path to graduation: Factors predicting on-time graduation rates,” *J. Coll. Student Retent. Res. Theory Pract.*, vol. 16, no. 3, pp. 351–371, 2014, doi: 10.2190/CS.16.3.c.
- [11] R. A. Pertiwi, Indahwati, F. M. Afendi, “Analisis Chaid Untuk Identifikasi Ketepatan Waktu Lulus Berdasarkan Karakteristik Mahasiswa,” *Xplore J. Stat.*, vol. 1, no. 2, 2013, doi: 10.29244/xplore.v1i2.12418.
- [12] Ilmu Komputer, “Buku Profil Program Studi Ilmu Teknologi Uin Sumatera Utara Medan Tahun 2017”, UIN Sumatera Utara Medan,

Medan, 2017.

- [13] D. Mulyana, “Metode Penelitian Kuantitatif,” *Metod. Penelit. Kualitatif*, p. 43, 2014.
- [14] Rudy Ansari, “Prediksi Kelulusan Mahasiswa Dengan Jaringan Syaraf Tiruan,” *J. Teknol. Inf. Univ. Lambung Mangkurat*, vol. 1, no. 1, pp. 18–23, 2016, doi: 10.20527/jtiulm.v1i1.4.
- [15] R. Rismala, S. Fatriandini, R. N. Dayawati, “Indonesia Symposium On Computing 2015: Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan Neuro-Fuzzy Informatika, Universitas Telkom” *Indones. Symp. Comput. 2015*, pp. 159–167, 2015.
- [16] N. I. Pratiwi, “Penggunaan Media Video Call dalam Teknologi Komunikasi,” *J. Ilm. Din. Sos.*, vol. 1, no. 2, pp. 202–224, 2017.
- [17] Han, M. Kamber, J. Pei, *Data mining Concept and Techniques 3rd edition*, Elsevier Inc., 2012
- [18] Syafnidawaty, “Metodologi Penelitian”,
<https://raharja.ac.id/2020/10/25/metodologi-penelitian/>
- [19] Depdikbud, "Pengertian Kerangka Penelitian, Macam Konsep, dan Contohnya", <https://dosensosiologi.com/kerangka-penelitian-pengertian-dan-contohnya-lengkap/>